ELSEVIER

Stochastic Processes and their Applications 92 (2001) 131–162

stochastic processes and their applications

www.elsevier.com/locate/spa

Large deviations for the Fleming–Viot process with neutral mutation and selection, II ☆

Donald A. Dawson[a, *], Shui Feng[b]

[a] The Fields Institute, 222 College Street, Toronto, ONT, Canada M5T 3J1

[b] Department of Mathematics and Statistics, McMaster University, Hamilton, ONT, Canada L8S 4K1

Received 5 January 2000; received in revised form 10 July 2000; accepted 6 August 2000

Abstract

Large deviation principles are established for the Fleming–Viot process with neutral mutation and with selection, and the associated equilibrium measures as the sampling rate approaches zero and when the state space is equipped with the weak topology. The path-level large deviation results improve the results of Dawson and Feng (1998, Stochastic Process. Appl. 77, 207–232) in three aspects: the state space is more natural, the initial condition is relaxed, and a large deviation principle is established for the Fleming–Viot process with selection. These improvements are achieved through a detailed study of the behaviour near the boundary of the Fleming–Viot process with finite types. © 2001 Elsevier Science B.V. All rights reserved.

MSC: primary 60F10; secondary 92D10

Keywords: Fleming–Viot process; Large deviations; Weak and τ-topology

1. Introduction

The Fleming–Viot process (henceforth, FV process) is a probability-valued stochastic process describing the evolution of the distribution of genotypes in a population under the influence of mutation, replacement sampling, and selective advantages among various genotypes.

Let $E=[0,1]$, $C(E)$ be the set of continuous functions on E, and $M_1(E)$ denote the space of all probability measures on E equipped with the usual weak topology and Prohorov metric ρ. Let A be the generator of a Markov process on E with domain $D(A)$. Define $\mathscr{D}=\{F : F(\mu)=f(\langle\mu,\phi\rangle),\ f\in C_{\mathrm{b}}^{\infty}(R), \phi\in D(A), \mu\in M_1(E)\}$, where $C_{\mathrm{b}}^{\infty}(R)$ denotes the set of all bounded, infinitely differentiable functions on R. Then the generator of the FV process in this article has the form

$$\mathscr{L}^{\gamma}F(\mu)=\int_E\left(A\frac{\delta F(\mu)}{\delta\mu(x)}\right)\mu(\mathrm{d}x)+\frac{\gamma}{2}\int_E\int_E\left(\frac{\delta^2F(\mu)}{\delta\mu(x)\delta\mu(y)}\right)Q(\mu;\mathrm{d}x,\mathrm{d}y)$$

☆ Research supported by the Natural Science and Engineering Research Council of Canada.

* Corresponding author.

E-mail addresses: don@fields.fields.utoronto.ca (D.A. Dawson), shuifeng@mcmail.cis.mcmaster.ca (S. Feng).

PII: S0304-4149(00)00070-3

$$=f'(\langle\mu,\phi\rangle)\langle\mu,A\phi\rangle+\frac{\gamma}{2}\int\int f''(\langle\mu,\phi\rangle)\phi(x)\phi(y)Q(\mu;\mathrm{d}x,\mathrm{d}y), \qquad (1)$$

where

$$\delta F(\mu)/\delta\mu(x)=\lim_{\varepsilon\to 0+}\varepsilon^{-1}\{F((1-\varepsilon)\mu+\varepsilon\delta_x)-F(\mu)\},$$

$$\delta^2 F(\mu)/\delta\mu(x)\delta\mu(y)=\lim_{\varepsilon_1\to 0+,\varepsilon_2\to 0+}(\varepsilon_1\varepsilon_2)^{-1}\{F((1-\varepsilon_1-\varepsilon_2)\mu+\varepsilon_1\delta_x+\varepsilon_2\delta_y)-F(\mu)\},$$

$$Q(\mu;\mathrm{d}x,\mathrm{d}y)=\mu(\mathrm{d}x)\delta_x(\mathrm{d}y)-\mu(\mathrm{d}x)\mu(\mathrm{d}y),$$

and δ_x stands for the Dirac measure at $x\in E$. The domain of $\mathscr{L}$ is $\mathscr{D}$. E is called the type space or the space of alleles, A is known as the mutation operator, and the last term describes the continuous sampling with sampling rate γ. If the mutation operator has the form

$$Af(x)=\frac{\theta}{2}\int(f(y)-f(x))v_0(\mathrm{d}y)$$

with $v_0\in M_1(E)$, we call the process a FV process with neutral mutation.

For any symmetric bounded measurable function $V(z,y)$ on $E^{\otimes 2}$, let

$$V(\mu)=\int_E\int_E V(z,y)\mu(\mathrm{d}z)\mu(\mathrm{d}y)$$

and

$$\left\langle\frac{\delta V(\mu)}{\delta\mu},\frac{\delta F}{\delta\mu}\right\rangle=\int_E\int_E\int_E\frac{\delta F}{\delta\mu}(z)[V(z,y)-V(y,w)]\,\mu(\mathrm{d}z)\mu(\mathrm{d}y)\mu(\mathrm{d}w).$$

Then the generator of a FV process with neutral mutation and selection takes the form

$$\mathscr{L}_V^{\gamma}F(\mu)=\mathscr{L}^{\gamma}F(\mu)+\left\langle\frac{\delta V(\mu)}{\delta\mu},\frac{\delta F}{\delta\mu}\right\rangle, \qquad (1.2)$$

where V is called the fitness function which is assumed to be continuous in the sequel. A nice survey on FV process and their properties can be found in Ethier and Kurtz (1993). In particular, it is shown in Ethier and Kurtz (1993) that the martingale problem associated with generators $\mathscr{L}$ and $\mathscr{L}_V$ are well-posed.

Let $T>0$ be fixed, and $C([0,T],M_1(E))$ denote the space of all $M_1(E)$-valued, continuous functions on $[0,T]$. For any μ in $M_1(E)$, let $P_\mu^{\theta,\gamma,v_0}$ and $P_\mu^{\theta,\gamma,V,v_0}$ be the laws of the FV process with neutral mutation and FV process with neutral mutation and selection, respectively. Π_{θ,γ,v_0} and $\Pi_{\theta,\gamma,v_0,V}$ will represent the corresponding equilibrium measures.

Let $\mathscr{X}(E)$ be the space of all finitely additive, non-negative, mass one measures on E, equipped with the smallest topology such that for all Borel subset B of E, $\mu(B)$ is continuous in μ. The σ-algebra $\mathscr{B}$ of space $\mathscr{X}(E)$ is the smallest σ-algebra such that for all Borel subset B of E, $\mu(B)$ is a measurable function of μ. It is clear that $M_1(E)$ is a strict subset of $\mathscr{X}(E)$. In Dawson and Feng (1998), large deviation principle (henceforth, LDP) is established for equilibrium measures on space $\mathscr{X}(E)$ and partial results are obtained for the path-level LPDs on a strange space under stronger topologies. In the present article we will first establish the LDPs for the equilibrium measures on space $M_1(E)$, and compare these with results obtained in Dawson and Feng (1998). Secondly, we establish the LDPs for the FV process with neutral mutation and with selection as $\gamma \to 0$ on space $C([0,T], M_1(E))$. These improve the corresponding results in Dawson and Feng (1998) in three aspects: the space is more natural, the initial condition is relaxed, and a full large deviation principle is established for the FV process with selection. This type of LDPs can be viewed as the infinite-dimensional generalization of the Freidlin–Wentzell theory. We prove the results through a detailed analysis of the boundary behaviour of the FV process with finite types.

In Section 2, we list some preliminary results to make the paper self-contained. LDPs for the equilibrium measures are the content of Section 3. The detailed study of the Fleming–Viot process with finite type space is carried out in Section 4. Finally, in Section 5, we prove the LDPs for the FV process with neutral mutation and with selection.

2. Preliminary

In this section, we give some definitions and results in the theory of large deviations. A more complete introduction to the theory is found in Dembo and Zeitouni (1993). Properties of finite additive measures will also be discussed.

Let X be a Hausdorff space with σ-algebra $\mathscr{F}$. Here $\mathscr{F}$ could be smaller than the Borel σ-algebra on X. $\{P_\gamma: \gamma > 0\}$ is a family of probability measures on $(X, \mathscr{F})$.

Definition 2.1. The family $\{P_\gamma: \gamma > 0\}$ is said to satisfy a LDP with a good rate function I if

1. for all $x \in X, I(x) \geqslant 0$;
2. for any $c \geqslant 0$, the level set $\Phi(c) = \{x \in X: I(x) \leqslant\}$ is compact in X;
3. for any $\mathscr{F}$-measurable closed subset F of X,
$$\limsup_{\gamma \to 0} \gamma \log P_\gamma(F) \leqslant -\inf_{x \in F} I(x);$$
4. for any $\mathscr{F}$-measurable open subset G of X,
$$\liminf_{\gamma \to} \gamma \log P_\gamma(G) \geqslant -\inf_{x \in G} I(x).$$

Functions satisfying the first two conditions are called good rate functions.

Definition 2.2. The family $\{P_\gamma: \gamma > 0\}$ is said to be exponentially tight if for any $\alpha > 0$ there exists a compact subset K_α of X such that
$$\limsup_{\gamma \to 0} \gamma \log P_\gamma(K_\alpha^{\mathrm{c}}) \leqslant -\alpha,$$
where K_α^{c} is the complement of K_α in X.

The following result of Pukhalskii will be used repeatedly in the sequel.

Theorem 2.1 (Pukhalskii). *Assume that the space X is a Polish space with metric d and $\mathscr{F}$ is the Borel σ-algebra. Then the family $\{P_\gamma: \gamma > 0\}$ satisfies a LDP with good rate function I if and only if the family $\{P_\gamma: \gamma > 0\}$ is exponentially tight, and for any $x \in X$*

$$\begin{aligned} -I(x) &= \lim_{\delta\to 0} \limsup_{\gamma\to 0} \gamma \log P_\gamma\{y \in X:\ \mathrm{d}(y,x) \leqslant \delta\} \\ &= \lim_{\delta\to 0} \liminf_{\gamma\to 0} \gamma \log P_\gamma\{y \in X:\ \mathrm{d}(y,x) < \delta\}. \end{aligned} \tag{2.3}$$

If the family $\{P_\gamma: \gamma > 0\}$ is exponentially tight, then for any sequence (γ_n) converging to zero from above as n goes to infinity, there is a subsequence (γ_{n_k}) of (γ_n) such that $\{P_{\gamma_{n_k}}\}$ satisfies a LDP with certain good rate function that depends on the subsequence.

Proof. Part one and part two are Corollary 3.4 and Theorem (P) of Pukhalskii (1991) respectively. $\square$

Let X be a Polish space, $M_1(X)$ denote the space of all probability measures on X equipped with the weak topology. Let $C_\mathrm{b}(X)$ and $B_\mathrm{b}(X)$ denote the set of bounded continuous functions on X, and the set of bounded measurable functions on X, respectively. For any μ, v in $M_1(X)$, the relative entropy of μ with respect to v is defined and denoted by

$$H(\mu|v) = \begin{cases} \displaystyle\int_X \phi \log \phi \, \mathrm{d}v & \text{if } \mu \ll v, \\ \infty & \text{otherwise,} \end{cases} \tag{2.4}$$

where ϕ is the Radon–Nikodym derivative of μ with respect to v. It is known (cf. Donsker and Varadhan, 1975) that

$$\begin{aligned} H(\mu|v) &= \sup_{g \in C_\mathrm{b}(X)} \left\{ \int_X g \, \mathrm{d}\mu - \log \int_X \mathrm{e}^g \mathrm{d}v \right\} \\ &= \sup_{g \in B_\mathrm{b}(X)} \left\{ \int_X g \, \mathrm{d}\mu - \log \int_X \mathrm{e}^g \, \mathrm{d}v \right\}. \end{aligned} \tag{2.5}$$

The relative entropy is closely related to the rate functions discussed in Section 3.

Finally, we present some results on finitely additive measures in Yosida and Hewitt (1952). Assume that the space X is a Polish space and $\mathscr{F}$ is the Borel σ-algebra. Let $\mathscr{N}(X)$ denote the set of all nonnegative, finitely additive measures on $(X, \mathscr{F})$. For any μ, v in $\mathscr{N}(X)$, we write $\mu \leqslant v$ if for any B in $\mathscr{F}$ we have $\mu(B) \leqslant v(B)$.

Definition 2.3. Let μ be any element in $\mathscr{N}(X)$. If 0 is the only countably additive measure v satisfying $0 \leqslant v \leqslant \mu$, then μ is called a pure finitely additive measure.

Here is an example of a pure finitely additive measure. Let $X = [0,1]$, $\mathscr{F}$ be the Borel σ-algebra. Define measure μ such that it only assumes the values of 0 and 1, $\mu(A) = 1$ if A contains $[0,a)$ for some $a > 0$, $\mu(A) = 0$ if A has Lebesgue measure zero. The existence of such a finitely additive measure is guaranteed by Theorem 4.1

in Yosida and Hewitt (1952). Since $\lim_{n\to\infty}\mu([0,1/n)) = 1 \neq 0 = \mu(\{0\})$, μ is not countably additive. To see that it is pure finitely additive, let v be any countably additive measure satisfying $0 \leqslant v \leqslant \mu$. Then

$$v([0,1]) = v(\{0\}) + \lim_{n\to\infty}(1 - v((0,1/n])) = 0$$

which implies that $v \equiv 0$.

Theorem 2.2 (Yosida and Hewitt, 1952, Theorem 1.19). *Let μ be any non-negative pure finitely additive measure, and v be any non-negative countably additive measure on space $(X,\mathscr{F})$. Then for any $\varepsilon > 0$, there exists A in $\mathscr{F}$ such that $\mu(A^c) = 0$, $v(A) < \varepsilon$, where A^c is the complement of A.*

Theorem 2.3. *Any non-negative measure μ on space $(X,\mathscr{F})$ can be uniquely written as the sum of a non-negative, countably additive measure μ_c and a non-negative, pure finitely additive measure μ_p.*

Proof. This is a combination of Theorems 1.23 and 1.24 in Yosida and Hewitt (1952). □

3. Equilibrium LDPs

Let $\mathscr{X}(E)$ be the space of all finitely additive, non-negative, mass one measures on E equipped with the projective limit topology, i.e., the weakest topology such that for all Borel subset B of E, $\mu(B)$ is continuous in μ. Under this topology, $\mathscr{X}(E)$ is Hausdorff. The σ-algebra $\mathscr{B}$ of space $\mathscr{X}(E)$ is the smallest σ-algebra such that for all Borel subset B of E, $\mu(B)$ is a measurable function of μ.

It was incorrectly stated in Dawson and Feng (1998) that $\mathscr{X}(\mathrm{E})$ can be identified with $M_1(E)$ equipped with the τ-topology for large deviation purposes. In this section we will first clarify the issues associated with equilibrium LDPs on the space $\mathscr{X}(E)$, and then establish the LDPs for equilibrium measures on $M_1(E)$ under the weak topology. Recall that the τ-topology on $M_1(E)$ is the smallest topology such that $\langle\mu, f\rangle = \int_E f(x)\,\mu(\mathrm{d}x)$ is continuous in μ for any bounded measurable function f on E. This topology is clearly stronger than the weak topology, and is the same as the subspace topology inherited from $\mathscr{X}(E)$. We use $M_1^\tau(E)$ to denote space $M_1(E)$ equipped with the τ-topology.

Theorem 3.1. *Every element μ in $\mathscr{X}(E)$ has the following unique decomposition*:

$$\mu = \mu_{ac} + \mu_s + \mu_p, \tag{3.1}$$

where μ_p is a pure finitely additive measure, μ_{ac} and μ_s are both countably additive with $\mu_{ac} \ll v_0$, $\mu_s \perp v_0$.

Proof. The result follows from Theorem 2.3 and the Lebesgue decomposition theorem. □

Remark. An element μ in $\mathscr{X}(E)$ is a probability measure if and only if $\mu_p(E) = 0$.

For any μ in $\mathscr{X}(E)$ satisfying $\mu_{ac}(E) > 0$, let $\mu_{ac}(\cdot|E) = \mu_{ac}(\cdot)/\mu_{ac}(E)$. Then it is clear that $\mu_{ac}(\cdot|E)$ is in $M_1(E)$. μ is said to be absolutely continuous with respect to v_0, still denoted by $\mu \ll v_0$, if $v_0(B) = 0$ implies $\mu(B) = 0$. For any two probability measures μ, v, $H(\mu|v)$ denotes the relative entropy of μ with respect to v. Define

$$I(\mu) = \begin{cases} \theta[H(v_0|\mu_{ac}(\cdot|E)) - \log \mu_{ac}(E)] & \text{if } \mu \ll v_0,\ \mu_{ac}(E) > 0, \mu \notin M_1(E), \\ \theta H(v_0|\mu) & \text{if } \mu \ll v_0,\ \mu \in M_1(E), \\ \infty & \text{else}. \end{cases} \tag{3.2}$$

Remark. $\mu \ll v_0$ implies that $\mu_s = 0$.

Theorem 3.2. *The family $\{\Pi_{\theta,\gamma,v_0}\}$ satisfies a LDP on $\mathscr{X}(E)$ with good rate function I.*

Proof. Let

$$\mathscr{P} = \{\{B_1, \ldots, B_r\}: r \geqslant 1, B_1, \ldots, B_r \text{ is a partition of } [0,1] \text{ by Borel measurable sets}\}. \tag{3.3}$$

Elements of $\mathscr{P}$ are denoted by $\imath, \jmath$, and so on. We say $\jmath \succ \imath$ iff $\jmath$ is finer than $\imath$. Then $\mathscr{P}$ partially ordered by $\succ$ is a partially ordered right-filtering set.

For every $\jmath = (B_1, \ldots, B_r) \in \mathscr{P}$, let

$$\mathscr{X}_\jmath = \left\{ x = (x_{B_1}, \ldots, x_{B_r}): x_{B_i} \geqslant 0,\ i = 1, \ldots, r;\ \sum_{i=1}^{r} x_{B_i} = 1 \right\}.$$

For any $\imath = (C_1, \ldots, C_l)$, $\jmath = (B_1, \ldots, B_r) \in \mathscr{P}$, $\jmath \succ \imath$, define

$$\pi_{\imath\jmath}: \mathscr{X}_\jmath \to \mathscr{X}_\imath, \quad (x_{B_1}, \ldots, x_{B_r}) \to \left(\sum_{B_k \subset C_1} x_{B_k}, \ldots, \sum_{B_k \subset C_l} x_{B_k} \right).$$

Then $\{\mathscr{X}_\jmath, \pi_{\imath\jmath}, \imath, \jmath \in J, \succ\}$ becomes a projective system, and the projective limit of this system can be identified as $\mathscr{X}(E)$.

For any finite partition $\imath = (B_1, \ldots, B_r)$ of E, and any μ in $\mathscr{X}(E)$, let

$$I_\imath(\mu) = \begin{cases} \sum_{k=1}^{r} v_0(B_k) \log \dfrac{v_0(B_k)}{\mu(B_k)} & \text{if } \mu \ll v_0, \\ \infty & \text{else}, \end{cases}$$

where we treat $c/0$ as infinity for $c > 0$.

Let

$$\tilde{I}(\mu) = \theta \sup_\imath I_\imath(\mu), \tag{3.4}$$

where the supremum is taken over all finite partitions of E.

By the standard monotone class argument indicator function over an interval can be approximated by bounded continuous functions pointwise. Hence, the restriction of the σ-algebra $\mathscr{B}$ on $M_1(E)$ coincides with the Borel σ-algebra generated by the weak topology, and Π_{θ,γ,v_0} is well-defined on space $(\mathscr{X}(E), \mathscr{B})$.

By using Theorem 3.3 of Dawson and Gärtner (1987), one gets that Π_{θ,γ,v_0} satistifies a LDP with good rate function $\tilde{I}$. Hence to prove the theorem it suffices to verify that

$I(\mu)=\tilde{I}(\mu)$. This is true if μ is not absolutely continuous with respect to v_0 since both are infinity. The case of μ in $M_1(E)$ follows from Lemma 2.3 in Dawson and Feng (1998). Now assume $\mu \ll v_0$ and $\mu \notin M_1(E)$. Then we have $\mu_{\mathrm{ac}}(E)<1, \mu_{\mathrm{p}} \neq 0$. If $\mu_{\mathrm{ac}}(E)=0$, then by applying Theorem 2.2, both I and $\tilde{I}$ are infinity. Next we assume that $\alpha=\mu_{\mathrm{ac}}(E)$ is in $(0,1)$. By definition we have $I(\mu) \geqslant I_\imath(\mu)$ for any finite partition $\imath$ of E. Thus $I \geqslant \tilde{I}$. On the other hand, for any $n \geqslant 1$ choose a set A_n such that $v_0(A_n)<1/n^2, \mu_{\mathrm{p}}(A_n^{\mathrm{c}})=0$. This is possible because Theorem 2.2 and μ_{p} is pure finitely additive. It is clear that $v_0(\bigcap_{i=1}^n A_i)<1/n^2, \mu_{\mathrm{p}}((\bigcap_{i=1}^n A_i)^{\mathrm{c}})=0$. Hence by taking intersection, the sequence $\{A_n\}$ can be chosen to be decreasing. For any finite partition $\imath=(B_1,\ldots,B_r)$ we introduce a new finite partition $\jmath=(B_1\cap A_n, B_1\cap A_n^{\mathrm{c}}, \ldots, B_r\cap A_n, B_r\cap A_n^{\mathrm{c}})$. Note that for any $p_i, x_i \geqslant 0$, $i=1,2$ we have the inequality

$$(p_1+p_2)\log\frac{p_1+p_2}{x_1+x_2} \leqslant p_1\log\frac{p_1}{x_1}+p_2\log\frac{p_2}{x_2}.$$

This implies

$$I_\imath(\mu) \leqslant I_\jmath(\mu)$$

and

$$\begin{aligned}
I_\jmath(\mu) &= \sum_{k=1}^r v_0(B_k\cap A_n^{\mathrm{c}})\log\frac{v_0(B_k\cap A_n^{\mathrm{c}})}{\mu(B_k\cap A_n^{\mathrm{c}})}+\sum_{k=1}^r v_0(B_k\cap A_n)\log\frac{v_0(B_k\cap A_n)}{\mu(B_k\cap A_n)}\\
&\geqslant \sum_{k=1}^r v_0(B_k\cap A_n^{\mathrm{c}})\log\frac{v_0(B_k\cap A_n^{\mathrm{c}})}{\mu(B_k\cap A_n^{\mathrm{c}})}+v_0(A_n)\log\frac{v_0(A_n)}{\mu(A_n)}\\
&= \sum_{k=1}^r v_0(B_k\cap A_n^{\mathrm{c}})\log\frac{v_0(B_k\cap A_n^{\mathrm{c}})}{\mu(B_k\cap A_n^{\mathrm{c}})}+v_0(A_n)\log\frac{v_0(A_n)}{\mu(A_n)\vee\alpha}.
\end{aligned}$$

Letting n go to infinity, we get

$$\begin{aligned}
I_\jmath(\mu) &\geqslant \lim_{n\to\infty}\sum_{k=1}^r v_0(B_k\cap A_n^{\mathrm{c}})\log\frac{v_0(B_k\cap A_n^{\mathrm{c}})}{\mu(B_k\cap A_n^{\mathrm{c}})}\\
&= \sum_{k=1}^r \lim_{n\to\infty} v_0(B_k\cap A_n^{\mathrm{c}})\log\frac{v_0(B_k\cap A_n^{\mathrm{c}})}{\mu(B_k\cap A_n^{\mathrm{c}})}\\
&\geqslant \sum_{k=1}^r v_0(B_k\cap F)\log\frac{v_0(B_k\cap F)}{\mu_{\mathrm{ac}}(B_k)},
\end{aligned}$$

where $F=\bigcup_n A_n^{\mathrm{c}}$. Since $v_0(F^{\mathrm{c}})=0$, we have

$$\tilde{I}(\mu) \geqslant \theta I_\jmath(\mu) \geqslant \theta\sum_{k=1}^r v_0(B_k)\log\frac{v_0(B_k)}{\mu_{\mathrm{ac}}(B_k)},$$

which implies that

$$\tilde{I}(\mu) \geqslant \theta\sup_\imath\left\{\sum_{k=1}^r v_0(B_k)\log\frac{v_0(B_k)}{\mu_{\mathrm{ac}}(B_k)}\right\}=I(\mu). \qquad \square$$

Lemma 3.3. *Assume that there is a sequence of decreasing intervals A_n such that the length of A_n converges to zero as n goes to infinity, $\nu_0(A_n)>0$ for all n, and $\bigcap_n A_n=\{x_0\}$ with $\nu_0(x_0)=0$. Then $I(\mu)$ defined above is not a good rate function on $M_1^\tau(E)$.*

Remark. Clearly a large class of probability measures including Lebesgue measure satisfy the condition in Lemma 3.3. But pure atomic measures with finite atoms do not satisfy the condition.

Proof. We will construct a counter example. Assume that $I(\mu)$ is a good rate function. Then for any $\beta>0$, the level set $\Phi(\beta)=\{\mu\in M_1(E)\colon I(\mu)\leqslant\beta\}$ is a τ-compact set. For any $\alpha\in(0,1)$, $n\geqslant 1$, choose $\mu_n(\mathrm{d}x)=f_n(x)\nu_0(\mathrm{d}x)$ with

$$f_n(x)=\frac{\alpha}{\nu_0(A_n)}\chi_{A_n}(x)+\frac{1-\alpha}{1-\nu_0(A_n)}\chi_{A_n^{\mathrm{c}}}(x),$$

where χ_A is the indicator function of set A, and A^{c} denotes the complement of A. By definition, we have

$$\begin{aligned}I(\mu_n)&=\theta H(\nu_0|\mu_n)=\theta\int_E\log\left(\frac{1}{f_n}\right)\nu_0(\mathrm{d}x)\\&=\theta\int_{A_n}\log\left(\frac{\nu_0(A_n)}{\alpha}\right)\nu_0(\mathrm{d}x)+\theta\int_{A_n^{\mathrm{c}}}\log\left(\frac{\nu_0(A_n^{\mathrm{c}})}{1-\alpha}\right)\nu_0(\mathrm{d}x)\\&\leqslant\log\frac{1}{\alpha(1-\alpha)},\end{aligned}$$

which implies that $\mu_n\in\Phi(\beta)$ with $\beta=\log 1/\alpha(1-\alpha)$. Since τ compactness implies the τ sequential compactness (see Gänssler, 1971, Theorem 2.6), the sequence μ_n converges in τ topology to a measure μ in $\Phi(\mu)$ and thus in weak topology to the same measure. For any continuous function g on E, one has

$$\begin{aligned}\lim_{n\to\infty}\langle\mu_n,g\rangle&=\lim_{n\to\infty}\left[\int_{A_n}\frac{\alpha g(x)}{\nu_0(A_n)}\nu_0(\mathrm{d}x)+\int_{A_n^{\mathrm{c}}}\frac{(1-\alpha)g(x)}{\nu_0(A_n^{\mathrm{c}})}\nu_0(\mathrm{d}x)\right]\\&=\alpha g(x_0)+\int_{\{x_0\}^{\mathrm{c}}}(1-\alpha)g(x)\nu_0(\mathrm{d}x),\end{aligned}$$

which implies that μ_n converges weakly to $\mu=\alpha\delta_{\{x_0\}}(\mathrm{d}x)+(1-\alpha)\nu_0(\mathrm{d}x)$. This lead to the contradiction

$$H(\nu_0|\mu)=\infty\leqslant\beta.\qquad\square$$

Lemma 3.4. *Assume ν_0 satisfies the condition in Lemma 3.3. Then it is impossible to establish a LDP for the Poisson–Dirichlet distribution with respect to ν_0 on $M_1^\tau(E)$ with a good rate function.*

Proof. Being the projective limit of a system of Hausdorff space, the space $\mathscr{X}(E)$ is also Hausdorff. By Tychonoff theorem, the product space $\Pi_{\iota\in\mathscr{P}}\mathscr{X}_\iota$ is compact. since $\mathscr{X}(E)$ is a closed subset of $\Pi_{\iota\in\mathscr{P}}\mathscr{X}_\iota$, it is also compact. Thus $\mathscr{X}(E)$ is a regular topological space. Assume that a LDP is true on $M_1^\tau(E)$ with a good rate function $J(\mu)$. By Lemma 3.3, J must differ from I. But the following arguments will lead to the equality of the two which is an obvious contradiction.

Fix an μ_0 in $M_1(E)$. The LDP on $M_1^\tau(E)$ with good J implies the corresponding LDP on $M_1(E)$ with good J. Since $M_1(E)$ is regular, we get that for any $\delta > 0$ there is an open neighborhood G^w of μ_0 such that

$$\inf_{\mu \in \overline{G^w}} J(\mu) \geqslant (J(\mu_0) - \delta) \wedge \frac{1}{\delta},$$

where $\overline{G^w}$ is the closure in space $M_1(E)$.

Let $G^{\mathscr{X}(E)}$ be an open set in $\mathscr{X}(E)$ such that $G^{\mathscr{X}(E)} \cap M_1(E) = G^w$. This is possible because the subspace topology on $M_1(E)$ inherited from $\mathscr{X}(E)$ is stronger than the weak topology. Now from the two LDPs, we get

$$-\inf_{G^w} I(\mu) \leqslant -\inf_{G^{\mathscr{X}}} I(\mu) \leqslant \liminf_{\gamma \to 0} \gamma \log \Pi_{\theta,\gamma,v_0}(G^{\mathscr{X}(E)}) = \liminf_{\gamma \to 0} \gamma \log \Pi_{\theta,\gamma,v_0}(G^w)$$

$$\leqslant \limsup_{\gamma \to 0} \gamma \log \Pi_{\theta,\gamma,v_0}(\overline{G^w}) \leqslant -\inf_{\overline{G^w}} J(\mu),$$

which implies that

$$I(\mu_0) \geqslant \inf_{G^w} I(\mu) \geqslant \inf_{\overline{G^w}} J(\mu) \geqslant (J(\mu_0) - \delta) \wedge \frac{1}{\delta}.$$

Letting δ approach zero we end up with $I(\mu_0) \geqslant J(\mu_0)$. On the other hand, since $\mathscr{X}(E)$ is also regular, we get that for any $\delta > 0$, there exists open set $G^{\mathscr{X}(E)}$ containing μ_0 such that

$$\inf_{\mu \in \overline{G^{\mathscr{X}(E)}}} I(\mu) \geqslant (I(\mu_0) - \delta) \wedge \frac{1}{\delta}$$

and $\overline{G^{\mathscr{X}(E)}}$ is the closure in $\mathscr{X}(E)$.

Let $G = G^{\mathscr{X}(E)} \cap M_1(E)$. Then as before we get

$$-\inf_{G} J(\mu) \leqslant \liminf_{\gamma \to 0} \gamma \log \Pi_{\theta,\gamma,v_0}(G) = \liminf_{\gamma \to 0} \gamma \log \Pi_{\theta,\gamma,v_0}(G^{\mathscr{X}(E)})$$

$$\leqslant \limsup_{\gamma \to 0} \gamma \log \Pi_{\theta,\gamma,v_0}(\overline{G^{\mathscr{X}(E)}}) \leqslant -\inf_{\overline{G^{\mathscr{X}(E)}}} I(\mu),$$

which implies that

$$J(\mu_0) \geqslant \inf_{G} J(\mu) \geqslant \inf_{\overline{G^{\mathscr{X}(E)}}} I(\mu) \geqslant (I(\mu_0) - \delta) \wedge \frac{1}{\delta}$$

and $J(\mu_0) \geqslant I(\mu_0)$. $\square$

From Theorem 3.1 and Lemma 3.4 we can see that in order to get an equilibrium LDP in the τ topology, one has to expand $M_1(E)$ to a bigger space. Next we are going to show that under a weaker topology, the weak topology, the equilibrium LDP holds on $M_1(E)$.

First note that the space $M_1(E)$ is a compact, Polish space with Prohorov metric ρ. Hence the sequence Π_{θ,γ,v_0} is exponentially tight. By Theorem 2.1, to obtain a LDP for $\Pi_{\theta,\gamma,\nu_0}$ with a good rate function it suffices to verify that there exists a function J such that for every $\nu \in M_1(E)$,

$$\lim_{\delta\to 0}\liminf_{\gamma\to 0}\gamma\log\Pi_{\theta,\gamma,\nu_0}\{\rho(\mu,\nu)<\delta\}$$
$$=\lim_{\delta\to 0}\limsup_{\gamma\to 0}\gamma\log\Pi_{\theta,\gamma,\nu_0}\{\rho(\mu,\nu)\leqslant\delta\}=-J(\nu). \tag{3.5}$$

By Theorem 2.1, the function J is the good rate function.

Let $\mathrm{supp}(\mu)$ denote the support of a probability measure μ, and $M_{1,\nu_0}(E)=\{\mu\in M_1(E)\colon \mathrm{supp}(\mu)\subset\mathrm{supp}(\nu_0)\}$. Let $\{\mu_n\}$ be an arbitrary sequence in $M_{1,\nu_0}(E)$ that converges to a μ in $M_1(E)$. Since $\mathrm{supp}(\nu_0)$ is a closed set, we get

$$1=\limsup_{n\to\infty}\mu_n\{\mathrm{supp}(\nu_0)\}\leqslant\mu\{\mathrm{supp}(\nu_0)\},$$

which implies that $\mu\in M_{1,\nu_0}(E)$. Hence $M_{1,\nu_0}(E)$ is a closed subset of $M_1(E)$. Next we prove (3.5) for

$$J(\nu)=\begin{cases}\theta H(\nu_0|\nu) & \text{if } \nu\in M_{1,\nu_0}(E),\\ \infty & \text{else,}\end{cases}$$

We will treat $0\log\frac{0}{0}$ as zero.

For any $\nu\in M_1(E)$, define $E_\nu=\{t\in(0,1)\colon \nu(\{t\})=0\}$. For any $t_1<t_2<\cdots<t_k\in E_\nu$, set $\nu^{t_1,\dots,t_k}=(\nu([0,t_1)),\dots,\nu([t_k,1]))$ which can be viewed as a probability measure on space $\{0,1,\dots,k\}$ with a probability $\nu([t_i,t_{i+1}))$ at $i\neq k$ and $\nu([t_k,1])$ at k. Set $t_0=0$.

Lemma 3.5. *For any* $\mu,\nu\in M_1(E)$,

$$H(\mu|\nu)=\sup_{t_1<t_2<\cdots<t_k\in E_\nu,k\geqslant 1}H(\mu^{t_1,\dots,t_k}|\nu^{t_1,\dots,t_k}). \tag{3.6}$$

Proof. By Lemma 2.3 in Dawson and Feng (1998), we have

$$H(\mu|\nu)\geqslant\sup_{t_1<t_2<\cdots<t_k\in E_\nu,k\geqslant 1}H(\mu^{t_1,\dots,t_k}|\nu^{t_1,\dots,t_k}). \tag{3.7}$$

On the other hand, by (2.5) for any $\varepsilon>0$, there is a continuous function g on E such that

$$H(\mu|\nu)\leqslant\int g\,\mathrm{d}\mu-\log\left(\int \mathrm{e}^g\,\mathrm{d}\nu\right)+\varepsilon.$$

Now choose $t_1^n<t_2^n\cdots<t_{k_n}^n$ in E_ν such that

$$\lim_{n\to\infty}\max_{i=0,\dots,k_n-1}\left[|t_{i+1}^n-t_i^n|+\max_{t,s\in[t_i^n,t_{i+1}^n]}|g(t)-g(s)|\right]=0.$$

This is possible because E_ν is a dense subset of E. Choose n large enough and let $t_0^n=0$, we get

$$H(\mu|\nu)\leqslant\sum_{i=0}^{k_n}g(t_i^n)\mu([t_i^n,t_{i+1}^n))+g(t_{k_n}^n)\mu([t_{k_n}^n,1])$$

$$-\log\left[\sum_{i=0}^{k_n-1} \mathrm{e}^{g(t_i^n)}\nu([t_i^n,t_{i+1}^n)) + \mathrm{e}^{g(t_{k_n}^n)}\nu([t_{k_n}^n,1])\right] + \varepsilon + \delta_n(g)$$

$$\leqslant \sup_{\alpha_i, i=0,\dots,k_n}\left\{\sum_{i=0}^{k_n} \alpha_i\mu([t_i^n,t_{i+1}^n)) + \alpha_{k_n}\mu([t_{k_n}^n,1])\right.$$

$$\left.-\log\left[\sum_{i=0}^{k_n-1} \mathrm{e}^{\alpha_i}\nu([t_i^n,t_{i+1}^n)) + \mathrm{e}^{\alpha_{k_n}}\nu([t_{k_n}^n,1])\right]\right\} + \varepsilon + \delta_n(g)$$

$$= H(\mu^{t_1^n,\dots,t_{k_n}}|\nu^{t_1^n,\dots,t_{k_n}}) + \varepsilon + \delta_n(g),$$

where $\delta_n(g)$ converges to zero as n goes to infinity. Letting n go to infinity, then ε go to zero, we get

$$H(\mu|\nu) \leqslant \sup_{t_1<t_2<\cdots<t_k \in E_\nu, k\geqslant 1} H(\mu^{t_1,\dots,t_k}|\nu^{t_1,\dots,t_k}).$$

This combined with (3.7) implies the result. □

For any $\delta > 0, \nu \in M_1(E)$, let

$$B(\nu,\delta) = \{\mu\in M_1(E)\colon \rho(\mu,\nu) < \delta\}, \quad \bar{B}(\nu,\delta) = \{\mu\in M_1(E)\colon \rho(\mu,\nu)\leqslant\delta\}.$$

Since the weak topology on $M_1(E)$ is generated by the family

$$\{\mu\in M_1(E)\colon f\in C_{\mathrm{b}}(E), x\in R, \varepsilon>0, |\langle\mu,f\rangle - x| < \varepsilon\},$$

there exist $f_1,\dots,f_m$ in $C_{\mathrm{b}}(E)$ and $\varepsilon > 0$ such that

$$\{\mu\in M_1(E)\colon |\langle\mu,f_j\rangle - \langle\nu,f_j\rangle| < \varepsilon\colon j=1,\dots,m\} \subset B(\nu,\delta).$$

Let

$$C = \sup\{|f_j(x)|\colon x\in E, j=1,\dots,m\},$$

and choose $t_1,\dots,t_k\in E_\nu$ such that

$$\sup\{|f_j(x)-f_j(y)|\colon x,y\in[t_i,t_{i+1}],\ i=0,1,\dots,k,\ t_{k+1}=1;\ j=1,\dots,m\} < \varepsilon/4.$$

Choosing $0<\delta_1<\varepsilon/2(k+1)C$, define

$$V_{t_1,\dots,t_k}(\nu,\delta_1) = \{\mu\in M_1(E)\colon |\mu([t_k,1]) - \nu([t_k,1])| < \delta_1,$$

$$|\mu([t_i,t_{i+1})) - \nu([t_i,t_{i+1}))| < \delta_1,\ i=0,\dots,k-1\}.$$

Then for any μ in $V_{t_1,\dots,t_k}(\nu,\delta_1)$ and any f_j, we have

$$|\langle\mu,f_j\rangle - \langle\nu,f_j\rangle| = \left|\int_{[t_k,1]} f_j(x)(\mu(\mathrm{d}x)-\nu(\mathrm{d}x))\right.$$

$$\left.+\sum_{i=0}^{k-1}\int_{[t_i,t_{i+1})} f_j(x)(\mu(\mathrm{d}x)-\nu(\mathrm{d}x))\right|$$

$$< \frac{\varepsilon}{2} + \sum_{i=0}^{k} |f_j(t_i)|\delta_1 < \varepsilon,$$

which implies that

$$V_{t_1,\dots,t_k}(v,\delta_1)\subset\{\mu\in M_1(E):\ |\langle\mu,f_j\rangle-\langle v,f_j\rangle|<\varepsilon:\ j=1,\dots,m\}\subset B(v,\delta).$$

Let

$$F(\mu)=(\mu([0,t_1)),\dots,\mu([t_k,1])).$$

Then $\Pi_{\theta,\gamma,v_0}\circ F^{-1}$ is a Dirichlet distribution with parameters $(\theta/\gamma)(\mu([0,t_1)),\dots,\mu([t_k,1]))$. By applying Theorem 2.2 in Dawson and Feng (1998) we get that for v in $M_{1,v_0}(E)$,

$$\begin{aligned}-J(v)&\leqslant-\theta H(v_0^{t_1,\dots,t_k}|v^{t_1,\dots,t_k})\\&\leqslant\liminf_{\gamma\to0}\ \gamma\log\Pi_{\theta,\gamma,v_0}\{V_{t_1,\dots,t_k}(v,\delta_1)\}\\&\leqslant\liminf_{\gamma\to0}\ \gamma\log\Pi_{\theta,\gamma,v_0}\{B(v,\delta)\}.\end{aligned}\tag{3.8}$$

Letting δ go to zero, we end up with

$$-J(v)\leqslant\lim_{\delta\to0}\ \liminf_{\gamma\to0}\ \gamma\log\Pi_{\theta,\gamma,v_0}\{B(v,\delta)\}.\tag{3.9}$$

For other v, (3.9) is trivially true.

On the other hand, for any $t_1,\dots,t_k$ in E_v, we claim that the vector function F is continuous at v. This is because all boundary points have v-measure zero. Hence for any $\delta_2>0$, there exists $\delta>0$ such that

$$\bar B(v,\delta)\subset V_{t_1,\dots,t_k}(v,\delta_2).$$

Let

$$\begin{aligned}\bar V_{t_1,\dots,t_k}(v,\delta_2)=\{\mu\in M_1(E):\ &|\mu([t_k,1])-v([t_k,1])|\leqslant\delta_1,\\&|\mu([t_i,t_{i+1}))-v([t_i,t_{i+1}))|\leqslant\delta_1,i=0,\dots,k-1\}.\end{aligned}$$

Then we have

$$\lim_{\delta\to0}\limsup_{\gamma\to0}\gamma\log\Pi_{\theta,\gamma,v_0}\{\bar B(v,\delta)\}\leqslant\lim_{\delta\to0}\limsup_{\gamma\to0}\gamma\log\Pi_{\theta,\gamma,v_0}\{\bar V_{t_1,\dots,t_k}(v,\delta_2)\}.\tag{3.10}$$

By letting δ_2 go to zero and applying Theorem 2.2 in Dawson and Feng (1998) to $\Pi_{\theta,\gamma,v_0}\circ F^{-1}$ again, one gets

$$\lim_{\delta\to0}\limsup_{\gamma\to0}\gamma\log\Pi_{\theta,\gamma,v_0}\{\bar B(v,\delta)\}\leqslant-J^{\theta}_{v_0^{t_1,\dots,t_k}}(v^{t_1,\dots,t_k}),\tag{3.11}$$

where

$$J^{\theta}_{v_0^{t_1,\dots,t_k}}(v^{t_1,\dots,t_k})=\begin{cases}\theta H(v_0^{t_1,\dots,t_k}|v^{t_1,\dots,t_k}) & \text{if } v^{t_1,\dots,t_k}\ll v_0^{t_1,\dots,t_k},\\ \infty & \text{else.}\end{cases}$$

Finally, taking supremum over the set E_v and applying Lemma 3.5, one gets

$$\lim_{\delta\to0}\limsup_{\gamma\to0}\gamma\log\Pi_{\theta,\gamma,v_0}\{\bar B(v,\delta)\}\leqslant-J(v),\tag{3.12}$$

which, combined with (3.9), implies the following theorem.

Theorem 3.6. *The family* $\{\Pi_{\theta,\gamma,\nu_0}\}$ *satisfies a LDP on* $M_1(E)$ *with good rate function* $J(\nu)$.

Remark. 1. By (3.2), for any μ in $M_1(E)$, we have that $I(\mu)=\infty$ if μ is not absolutely continuous with respect to ν_0. On the other hand, if we choose $\mu=\frac{1}{2}\nu_0+\frac{1}{2}\delta_0$, then μ is not absolutely continuous with respect to ν_0 but $J(\mu)<\infty$. Hence the restriction of $I(\nu)$ on $M_1(E)$ is not equal to $J(\nu)$.

2. Let $(\rho_1,\rho_2,\ldots)$ be a probability-valued random variable that has the Poisson–Dirichlet distribution with parameter θ/γ (cf. Kingman, 1975), and $\xi_1,\xi_2,\ldots,$ be i.i.d. with common distribution ν_0. Then $\Pi_{\theta,\gamma,\nu_0}$ is the distribution of $\sum_{i=1}^{\infty}\rho_i\delta_{\xi_i}$ (cf. Ethier and Kurtz, 1994, Lemma 4.2), and the LDP we obtained describes the large deviations in the following law of large numbers:

$$\sum_{i=1}^{\infty}\rho_i\delta_{\xi_i}\Rightarrow\nu_0.$$

The new features are clearly seen by comparing this with the Sanov theorem that describes the large deviations in the law of large numbers:

$$\sum_{i=1}^{n}\frac{1}{n}\delta_{\xi_i}\Rightarrow\nu_0.$$

Corollary 3.1. *The family* $\{\Pi_{\theta,\gamma,\nu_0,V}\}$ *satisfies a LDP on space* $M_1([0,1])$ *with good rate function* $J_V(\nu)=\sup_\mu\{V(\mu)-J(\mu)\}-(V(\nu)-J(\nu))$.

Proof. By Lemma 4.2 of Ethier and Kurtz (1994), one has

$$\Pi_{\theta,\gamma,\nu_0,V}(\mathrm{d}\mu)=Z^{-1}\exp\left[\frac{V(\mu)}{\gamma}\right]\Pi_{\theta,\gamma,\nu_0}(\mathrm{d}\mu), \tag{3.13}$$

where Z is the normalizing constant.

Since $V(x)$ is continuous, we get that $V(\mu)\in C(M_1([0,1]))$. By using Varadhan's Lemma, we have

$$\lim_{\gamma\to 0}\gamma\log Z=\lim_{\gamma\to 0}\gamma\log\int \mathrm{e}^{V(\mu)/\gamma}\Pi_{\theta,\gamma,\nu_0}(\mathrm{d}\mu)=\sup_\mu\{V(\mu)-J(\mu)\}. \tag{3.14}$$

By direct calculation, we get that for any $\nu\in M_1(E)$

$$\begin{aligned}
&\lim_{\delta\to 0}\liminf_{\gamma\to 0}\gamma\log\int_{B(\nu,\delta)}\mathrm{e}^{V(\mu)/\gamma}\Pi_{\theta,\gamma,\nu_0}(\mathrm{d}\mu)\\
&\quad=V(\nu)+\lim_{\delta\to 0}\liminf_{\gamma\to 0}\gamma\log\Pi_{\theta,\gamma,\nu_0}\{B(\nu,\delta)\}\\
&\quad=V(\nu)+\lim_{\delta\to 0}\limsup_{\gamma\to 0}\gamma\log\Pi_{\theta,\gamma,\nu_0}\{\bar{B}(\nu,\delta)\}\\
&\quad=\lim_{\delta\to 0}\liminf_{\gamma\to 0}\gamma\log\int_{\bar{B}(\nu,\delta)}\mathrm{e}^{V(\mu)/\gamma}\Pi_{\theta,\gamma,\nu_0}(\mathrm{d}\mu)\\
&\quad=V(\nu)-J(\nu),
\end{aligned}$$

which combined with (3.14) implies

$$\lim_{\delta\to 0}\liminf_{\gamma\to 0}\gamma\log\Pi_{\theta,\gamma,v_0,V}\{\rho(\mu,v)<\delta\}$$
$$=\lim_{\delta\to 0}\limsup_{\gamma\to 0}\gamma\log\Pi_{\theta,\gamma,v_0,V}\{\rho(\mu,v)\leqslant\delta\}=-J_V(v). \tag{3.15}$$

Since the family $\{\Pi_{\theta,\gamma,v_0,V}\}$ is also exponentially tight, using Theorem 2.1 again, we get the result. □

4. LDP for FV process with finite types

We now turn to the study of LDP at the path-level. In this section, we focus on FV process with finite types.

Let $E_n=\{1,2,\ldots,n\}$, and define

$$\mathscr{S}_n=\left\{x=(x_1,\ldots,x_{n-1})\colon\ x_i\geqslant 0,\ i=1,\ldots,n-1;\ \sum_{i=1}^{n-1}x_i\leqslant 1\right\},$$

with $\mathscr{S}_n^\circ$ denoting its interior. Then the FV process with n types is a FV process with neutral mutation and type space E_n. It is a finite-dimensional diffusion process statisfying the following system of stochastic differential equations:

$$\mathrm{d}x_k^\gamma(t)=b_k(x^\gamma(t))\mathrm{d}t+\sqrt{\gamma}\sum_{l=1}^{n-1}\sigma_{kl}(x^\gamma(t))\mathrm{d}B_l(t),\quad 1\leqslant k\leqslant n-1, \tag{4.1}$$

where $x^\gamma(t)=(x_1^\gamma(t),\ldots,x_{n-1}^\gamma(t))$, $b_k(x^\gamma(t))=\theta/2(p_k-x_k^\gamma(t))$, and $\sigma(x^\gamma(t))=(\sigma_{kl}(x^\gamma(t)))_{1\leqslant k,l\leqslant n-1}$ is given by

$$\sigma(x^\gamma(t))\sigma'(x^\gamma(t))=D(x^\gamma(t))=(x_k^\gamma(t)(\delta_{kl}-x_l^\gamma(t)))_{1\leqslant k,l\leqslant n-1},$$

where $p_k=v_0(k)$, and $B_l(t)$, $1\leqslant l\leqslant n-1$ are independent Brownian motions.

Let P_x^γ denote the law of $x^\gamma(\cdot)$ starting at x. The LDP for P_x^γ on space $C([0,T],\mathscr{S}_n)$ as γ goes to zero has been studied in Dawson and Feng (1998) under the assumptions that p_k is strictly positive for all k and x is in the interior of $\mathscr{S}_n$. In this section we will consider the LDP for P_x^γ when these assumptions are not satisfied, i.e., some of p_k are zero or x is on the boundary. This creates serious difficulties because of the degeneracy and the non-Lipschitz behaviour of the square root of the diffusion coefficient on the boundary. Let $p_n=1-\sum_{i=1}^{n-1}p_i$, $x_n=1-\sum_{i=1}^{n-1}x_i$, and $b_n(x)=(\theta/2)(p_n-x_n)$. If for a particular k, $p_k=x_k\in\{0,1\}$, then $x_k(t)$ will be zero (or one) for all positive t. Thus, without the loss of generality, we assume that p_k+x_k is not zero or two for all k. For any given $p=(p_1,\ldots,p_{n-1})$, define

$$\mathscr{L}_p=\{x=(x_1,\ldots,x_{n-1})\in\mathscr{S}_n\colon\ 0<x_k+p_k<2,k=1,\ldots,n\}.$$

For any x in $\mathscr{L}_p$, let $H_x^{\alpha,\beta}$ is the set of all absolutely continuous element in $C([\alpha,\beta],\mathscr{S}_n)$ starting at x, i.e.,

$$H_x^{\alpha,\beta}=\left\{\varphi\in C([\alpha,\beta],\mathscr{S}_n)\colon\ \varphi(t)=x+\int_\alpha^t\dot\varphi(s)\,\mathrm{d}s\right\}.$$

Define

$$I_x^{\alpha,\beta}(\varphi) = \begin{cases} \frac{1}{2}\int_\alpha^\beta \sum_{i=1}^n \frac{(\dot{\varphi}_i(t) - b_i(\varphi(t)))^2}{\varphi_i(t)}\,\mathrm{d}t, & \varphi \in H_x^{\alpha,\beta}, \\ \infty, & \varphi \notin H_x^{\alpha,\beta}, \end{cases} \tag{4.2}$$

where $\varphi_n(t) = 1 - \sum_{i=1}^{n-1}\varphi_i(t)$, $0/0 = 0$, $c/0 = \infty$ for $c > 0$, and the integrations are the Lebesgue integrals. We denote $I_x^{0,T}(H_x^{0,T})$ by I_x (resp. H_x).

Lemma 4.1. *For $n = 2$ and any $\varphi(\cdot)$ in $C([0,T],\mathscr{S}_2)$, we have*

$$\limsup_{\delta\to 0}\limsup_{\gamma\to 0}\gamma\log P_x^\gamma\left\{\sup_{t\in[0,T]}|x(t)-\varphi(t)|\leqslant\delta\right\} \leqslant -I_x(\varphi). \tag{4.3}$$

Proof. Since $n = 2$, we have $x = x_1$, $p = p_1$. The result has been proved in the case of $0 < x < 1$, $0 < p < 1$ in Theorem 3.3 of Dawson and Feng (1998). Next we consider the remaining cases: (A) $x = 0$, $0 < p < 1$; (B) $0 \leqslant x < 1, p = 1$; (C) $x = 1$, $0 < p < 1$; (D) $0 < x \leqslant 1, p = 0$. By dealing with $1 - x$ and $1 - p$, we can derive (C) and (D) from (A) and (B), respectively.

For any $\delta > 0, N \geqslant 1$, and $0 \leqslant a \leqslant b \leqslant T$, set

$$B(\varphi,\delta;a,b) = \left\{\psi \in C([0,T],\mathscr{S}_n)\colon \sup_{a\leqslant t\leqslant b}|\psi(t)-\varphi(t)|\leqslant\delta\right\},$$

$$B^\circ(\varphi,\delta;a,b) = \left\{\psi \in C([0,T],\mathscr{S}_n)\colon \sup_{a\leqslant t\leqslant b}|\psi(t)-\varphi(t)| < \delta\right\}.$$

If φ is not in H_x, either $\varphi(0) \neq x$ or φ is not absolutely continuous. Clearly the result holds in the case of $\varphi(0) \neq x$. If φ is not absolutely continuous, then there exist $c > 0$ and disjoint subintervals $[a_1^m, b_1^m], \ldots, [a_{k_m}^m, b_{k_m}^m]$ such that $\sum_{l=1}^{k_m}(b_l^m - a_l^m) \to 0$, while $\sum_{l=1}^{k_m}|\varphi(b_l^m) - \varphi(a_l^m)| \geqslant c$. By Chebyshev's inequality and martingale property, we get

$$\begin{aligned}
&\limsup_{\delta\to 0}\limsup_{\gamma\to 0}\gamma\log P_x^\gamma\{B(\varphi,\delta;0,T)\} \\
&\quad\leqslant \limsup_{\delta\to 0}\limsup_{\gamma\to 0}\gamma\log\left(E^{P_x^\gamma}\left\{\exp\left(\sum_{l=1}^{k_m}\frac{1}{\gamma}\left[\theta_l(x(b_l^m)-x(a_l^m))\right.\right.\right.\right. \\
&\qquad\left.\left.\left.\left. - \int_{a_l^m}^{b_l^m}\left(\theta_l b(x(s)) - \frac{\theta_l^2}{2}x(s)(1-x(s))\,\mathrm{d}s\right)\right]\right)\right\}\right. \\
&\qquad\times\inf_{\psi\in B(\varphi,\delta;0,T)}\exp\left(-\sum_{l=1}^{k_m}\frac{1}{\gamma}\left[\theta_l(\psi(b_l^m)-\psi(a_l^m)) - \int_{a_l^m}^{b_l^m}\left(\theta_l b(\psi(s))\right.\right.\right. \\
&\qquad\left.\left.\left.\left. -\frac{\theta_l^2}{2}\psi(s)(1-\psi(s))\,\mathrm{d}s\right)\right]\right)\right)
\end{aligned}$$

$$= -\sum_{l=1}^{k_m}\left[\theta_l(\varphi(b_l^m)-\varphi(a_l^m)) - \int_{a_l^m}^{b_l^m}(\theta_l b(\varphi(s)) - \frac{\theta_l^2}{2}\varphi(s)(1-\varphi(s))\,\mathrm{d}s\right]$$

$$\leqslant -c\kappa + C(\kappa)\sum_{l=1}^{k_m}(b_l^m - a_l^m),$$

where $\theta_l = \kappa\,\mathrm{sign}(\varphi(b_l^m)-\varphi(a_l^m))$, $l=0,\dots,k_m-1$, and $C(\kappa)$ is a positive constant depending on κ. Here

$$\mathrm{sign}(c) = \begin{cases} 1, & c>0,\\ -1, & c<0,\\ 0, & c=0. \end{cases}$$

Now let m go to infinity, and then let κ go to infinity, one ends up with

$$\limsup_{\delta\to 0}\limsup_{\gamma\to 0}\gamma\log P_x^\gamma\{B(\varphi,\delta;0,T)\} \leqslant -\infty = -I_x(\varphi). \tag{4.4}$$

Next we assume that $\varphi\in H_x$.

(A) $x=0, 0<p<1$. Four cases need to be treated separately based on the behaviour of φ.

Case I: There is a $t_0>0$ such that $\varphi(t)=0$ for all t in $[0,t_0]$.

In this case $I_x(\varphi)=\infty$. On the other hand, choose δ small enough such that

$$\delta < \min\left\{\frac{p}{2}, \frac{\theta p t_0}{8}\right\}.$$

Then for any $x(\cdot)$ in $C([0,T],\mathscr{S}_n)$ satisfying $\sup_{t\in[0,t_0]}x(t)\leqslant\delta$, we have

$$\sup_{t\in[0,t_0]}\left[\int_0^t b(x(s))\,\mathrm{d}s - x(t)\right] \geqslant \frac{\theta p t_0}{8}.$$

Next by choosing $c=\theta p/8\gamma\delta$, we get

$$\sup_{t\in[0,t_0]}\left[(-c)(x(t)-\int_0^t b(x(s))\,\mathrm{d}s) - \frac{\gamma c^2}{2}\int_0^t x(s)(1-x(s))\,\mathrm{d}s\right] \geqslant \frac{\theta^2 p^2 t_0}{128\gamma\delta},$$

which combined with Doob's inequality implies

$$P_x^\gamma\left\{\sup_{0\leqslant t\leqslant T}|x(t)-\varphi(t)|\leqslant\delta\right\} \leqslant P_x^\gamma\left\{\sup_{0\leqslant t\leqslant t_0}x(t)\leqslant\delta\right\}$$

$$\leqslant P_x^\gamma\left\{\sup_{0\leqslant t\leqslant t_0}\left[(-\rho)\left(x(t)-\int_0^t b(x(s))\,\mathrm{d}s\right)\right.\right.$$

$$\left.\left. -\frac{\rho^2\gamma}{2}\int_0^t x(s)(1-x(s))\,\mathrm{d}s\right] > \frac{\theta^2p^2t_0}{128\gamma\delta}\right\}$$

$$\leqslant \exp\left[-\left(\frac{\theta^2p^2t_0}{128\gamma\delta}\right)\right].$$

Letting γ go to zero, then δ go to zero, we get (4.3).

Case II: For all t in $(0,T]$, $\varphi(t)$ stays away from 0 and 1.

For any $N \geqslant 1$, choose δ small enough such that no functions in the set $B(\varphi, 2\delta; 1/N, T)$ hit zero or one in the time interval $[1/N, T]$. Let μ be the law of $x^{\gamma}_{1/N}$ under P^{γ}_x. Then one gets

$$\begin{aligned}
\gamma \log P^{\gamma}_x\{B(\varphi,\delta;0,T)\} &\leqslant \gamma \log P^{\gamma}_x\{B(\varphi,\delta;1/N,T)\}\\
&= \gamma \log \int_0^1 P^{\gamma}_y\{B(\varphi,\delta;1/N,T)\}\,\mu(\mathrm{d}y)\\
&\leqslant \gamma \log \sup_{|y-\varphi(1/N)|\leqslant\delta} P^{\gamma}_y\{B(\varphi,\delta;1/N,T)\}\\
&= \gamma \log P^{\gamma}_{y^{\gamma}}\{B(\varphi,\delta;1/N,T)\} \quad \text{for some } |y^{\gamma}-\varphi(1/N)|\leqslant\delta,
\end{aligned}$$

where in the last equality we used the property that the supremum of an upper semicontinuous function over a closed set can be reached at certain point inside the set. Noting that $P^{\gamma}_{y^{\gamma}}$ coincides with a non-degenerate diffusion over the interval $[1/N, T]$ on any set that does not hit the boundary of $[0, 1]$. By the uniform large deviation principle for non-degenerate diffusions (cf. Dembo and Zeitouni, 1993), we get

$$\limsup_{\gamma\to 0} \gamma \log P^{\gamma}_x\{B(\varphi,\delta;0,T)\} \leqslant -\inf_{|y-\varphi(1/N)|\leqslant\delta}\ \inf_{\psi\in B(\varphi,\delta;1/N,T)} I^{1/N,T}_y(\psi). \tag{4.5}$$

Assume that $\inf_{|y-\varphi(1/N)|\leqslant\delta}\inf_{\psi\in B(\varphi,\delta;1/N,T)} I^{1/N,T}_y(\psi)$ is finite for small δ. Otherwise $I_x(\varphi) \geqslant I^{1/N,T}_{\varphi(1/N)}(\varphi) = \infty$, and the upper bound is trivially true. For any y satisfying $|y-\varphi(1/N)|\leqslant\delta$, ψ in $B(\varphi,\delta;1/N,T)$ satisfying $\psi(1/N) = y$, $I^{1/N,T}_y(\psi) < \infty$, we define for t in $[1/N, T]$

$$\bar{\psi}(t) = \psi(t) + (\varphi(1/N) - y).$$

Then it is clear that $\bar{\psi}$ is in $B(\varphi, 2\delta; 1/N, T)$ and thus does not hit zero or one. By direct calculation, we get that

$$I^{1/N,T}_{\varphi(1/N)}(\bar{\psi}) \leqslant I^{1/N,T}_y(\psi) + \delta_N,$$

where δ_N goes to zero as δ goes to zero for any fixed N. This combined with (4.5) implies that

$$\begin{aligned}
&\limsup_{\delta\to 0}\ \limsup_{\gamma\to 0} \gamma \log P^{\gamma}_x\left\{\sup_{0\leqslant t\leqslant T}|x(t)-\varphi(t)|\leqslant\delta\right\}\\
&\leqslant -\lim_{\delta\to 0}\ \inf_{\psi\in B(\varphi,2\delta;1/N,T)} I^{1/N,T}_{\varphi(1/N)}(\psi) = -I^{1/N,T}_{\varphi(1/N)}(\varphi),
\end{aligned} \tag{4.6}$$

where the equality follows from the lower semicontinuity of $I^{1/N,T}_{\varphi(1/N)}(\cdot)$ at non-degenerate paths. Finally by letting N go to infinity we end up with (4.3).

More generally, if $\varphi(t)$ is in $(0, 1)$ over $[\alpha, \beta] \subset [0, 1]$, then the above argument leads to

$$\limsup_{\delta\to 0}\ \limsup_{\gamma\to 0} \gamma \log P^{\gamma}_x\left\{\sup_{0\leqslant t\leqslant T}|x(t)-\varphi(t)|\leqslant\delta\right\} \leqslant -I^{\alpha,\beta}_{\varphi(\alpha)}(\varphi). \tag{4.7}$$

Case III: $\varphi(t)$ is in $(0, 1]$ for all t in $(0, T]$, i.e., φ may hit boundary point 1.

Let

$$\tau_1 = \inf\{t \in [0,T]: \varphi(t) = 1\}, \qquad \tau = \inf\{t \in [0,T]: I_x^{0,t}(\varphi) = \infty\}.$$

If $\tau < \tau_1$, (4.3) is proved by applying the arguments in Case II to the time interval $[0, (\tau+\tau_1)/2]$. Now assume that $\tau \geqslant \tau_1$.

Since $p < 1$, one can find $0 < t_3 < t_2 < \tau_1$ satisfying $\inf_{s\in[t_3,\tau_1]}\varphi(s) > p$. By using result in Case II, we have

$$\limsup_{\gamma\to 0} \gamma \log P_x^\gamma \left\{ \sup_{t\in[0,T]} |x(t)-\varphi(t)| \leqslant \delta \right\} \leqslant -I_x^{0,t_2}(\varphi).$$

Since $I_x^{0,t_2}(\varphi)$ is finite for all t_2, we get

$$\begin{aligned} I_x(\varphi) \geqslant I_x^{0,t_2}(\varphi) &= \frac{1}{2}\int_0^{t_2} \frac{(\dot\varphi(t) - (\theta/2)(p-\varphi(t)))^2}{\varphi(t)(1-\varphi(t))}\,\mathrm{d}t \\ &\geqslant \frac{\theta}{2}\int_{t_3}^{t_2} (\varphi(t)-p)\frac{\dot\varphi(t)}{1-\varphi(t)}\,\mathrm{d}t \to \infty \quad \text{as } t_2 \nearrow \tau_1, \end{aligned} \tag{4.8}$$

which implies (4.3).

Case IV: A second visit to zero by $\varphi(t)$ occurs at a strictly positive time.

Let

$$\tau_0 = \inf\{t > 0: \varphi(t) = 0\} > 0.$$

Choosing $0 < t_1 < t_2 < \tau_0$ such that $\inf_{t\in[t_1,\tau_0]}(p - \varphi(t)) > 0$. Then we have

$$\begin{aligned} I_x(\varphi) &\geqslant \lim_{t_2\nearrow\tau_0} \frac{1}{2}\int_{t_1}^{t_2} \frac{(\dot\varphi(t)-(\theta/2)(p-\varphi(t)))^2}{\varphi(t)}\,\mathrm{d}t \\ &\geqslant -\lim_{t_2\nearrow\tau_0} \frac{\theta}{2}\int_{t_1}^{t_2}(p-\varphi(t))\frac{\dot\varphi(t)}{\varphi(t)}\,\mathrm{d}t = \infty. \end{aligned}$$

This combined with (4.7) implies the result.

(B) $0 \leqslant x < 1$, $p = 1$. First assume that $I_x(\varphi)$ is finite (i.e. $\tau = \infty$). For small ε, define

$$\Gamma_\varepsilon(\varphi) = \{t \in [0,T]: \varphi(t) < 1-\varepsilon\} = \bigcup_{i=1}^{\infty}(a_i, b_i).$$

Since the set $\{t\in[0,T]: (\dot\varphi(t) - (\theta/2)(1-\varphi(t)))^2/\varphi(t)(1-\varphi(t)) = 0\}$ has no contribution to the value of $I_x(\varphi)$ and the finiteness of $I_x(\varphi)$ implies that the set $\mathcal{N} = \{t\in[0,T]: (\dot\varphi(t)-(\theta/2)(1-\varphi(t)))^2/\varphi(t)(1-\varphi(t)) = \infty\}$ has zero Lebesgue measure, we may redefine the value of $(\dot\varphi(t)-(\theta/2)(1-\varphi(t)))^2/\varphi(t)(1-\varphi(t))$ to be zero on $\mathcal{N}$ without changing the value of the rate function. After this modification we can apply the monotone convergence theorem and get that $I_x^{\Gamma_\varepsilon}(\varphi)$ converges to $I_x(\varphi)$ as ε goes to zero.

By the Markov property and an argument similar to that used in deriving (4.5) and (4.6), we have for any $m \geqslant 1$ and $\varepsilon > 0$,

$$\limsup_{\delta\to 0}\limsup_{\gamma\to 0} \gamma \log P_x^\gamma \left\{ \sup_{t\in[0,T]} |x(t)-\varphi(t)| \leqslant \delta \right\}$$

$$\leqslant \sum_{i=1}^{m} \limsup_{\delta\to 0} \limsup_{\gamma\to 0} \gamma \sup_{|y-\varphi(a_i)|\leqslant\delta} \log P_x^{\gamma}\left\{ \sup_{t\in[a_i,b_i]} |x(t)-\varphi(t)|\leqslant\delta \right\}$$

$$\leqslant -\sum_{i=1}^{m} \liminf_{\delta\to 0} \inf_{|y-\varphi(a_i)|\leqslant\delta} \inf_{\psi\in B(\varphi,\delta;a_i,b_i)} I_y^{a_i,b_i}(\psi)$$

$$= -\sum_{i=1}^{m} I_{\varphi(a_i)}^{a_i,b_i}(\varphi). \tag{4.9}$$

By letting m go to infinity, and then ε go to zero, we get (4.3).

Next we assume the rate function is infinity. By an argument similar to that used in Case IV, we get the result for all paths that hit zero at a positive time. We now assume that φ does not hit zero at any positive time. If $\tau < \tau_1$ the result is true by using the argument in Case III. Let us now assume that $\tau_1 \leqslant \tau \leqslant T$.

If $\varphi(t)=1$ over $[\tau,\tau+\varepsilon]$ for some $\varepsilon>0$, then we have $\lim_{t\nearrow\tau} I_x^{0,t}(\varphi)=\infty$ and the result follows by approaching τ from below.

If $\varphi(\tau)$ is in $(0,1)$, then the result is obtained by using (4.7) in a small two-sided neighborhood of τ since the rate function over the neighborhood is infinity.

The only possibility left is that $\varphi(\tau)=1$, and $0<\varphi(t)<1$ over $(\tau,\beta]$ for some $\tau<\beta\leqslant T$. By applying (4.7) over $[\alpha,\beta]$ with $\alpha\in(\tau,\beta)$ and letting α approach τ from above, we get the result. □

Lemma 4.2. *For any $n\geqslant 2$ and any $\varphi(\cdot)$ in $C([0,T],\mathscr{S}_n)$, we have that*

$$\text{for any } x\in\mathscr{Z}_p,\ \limsup_{\delta\to 0} \limsup_{\gamma\to 0} \gamma \log P_x^{\gamma}\left\{ \sup_{t\in[0,T]} |x(t)-\varphi(t)|\leqslant\delta \right\}$$

$$\leqslant -I_x(\varphi), \tag{4.10}$$

$$\text{for any } x\in\mathscr{S}_n^{\circ}\cap\mathscr{Z}_p,\ \liminf_{\delta\to 0} \liminf_{\gamma\to 0} \gamma \log P_x^{\gamma}\left\{ \sup_{t\in[0,T]} |x(t)-\varphi(t)|<\delta \right\}$$

$$\geqslant -I_x(\varphi). \tag{4.11}$$

Proof. If $\varphi(t)\in C([0,T],\mathscr{S}_n)$ and $I_{x_i}(\varphi_i)=\infty$ for some $i=1,\ldots,n$, where $I_{x_i}(\varphi_i)$ represents the rate function for the two-type process $(x_i(\cdot),\sum_{j\neq i} x_j(\cdot))$, then we have

$$\limsup_{\delta\to 0} \limsup_{\gamma\to 0} \gamma \log P_x^{\gamma}\left\{ \sup_{t\in[0,T]} |x(t)-\varphi(t)|\leqslant\delta \right\}$$

$$\leqslant \limsup_{\delta\to 0} \limsup_{\gamma\to 0} \gamma \log P_x^{\gamma}\left\{ \sup_{t\in[0,T]} |x_i(t)-\varphi_i(t)|\leqslant\delta \right\}$$

$$\leqslant -I_{x_i}(\varphi_i) = -\infty. \tag{4.12}$$

Since for any k,l

$$\frac{(\dot\varphi_k(t)+\dot\varphi_l(t)-b_k(\varphi(t))-b_l(\varphi(t)))^2}{\varphi_k(t)+\varphi_l(t)}$$

$$\leqslant \frac{(\dot\varphi_k(t) - b_k(\varphi(t)))^2}{\varphi_k(t)} + \frac{(\dot\varphi_l(t) - b_l(\varphi(t)))^2}{\varphi_l(t)},$$

we conclude that $I_x(\varphi) = \infty$ implies that $I_{x_k}(\varphi_k) = \infty$ for some $1 \leqslant k \leqslant n$. Without the loss of generality, let us assume that $k = 1$. Thus by (4.12) and by applying Lemma 4.1 to the two-type process $(x_1(\cdot), \sum_{j=2}^n x_j(\cdot))$ we get both (4.10) and (4.11) when $I_x(\varphi) = \infty$.

Assume that $I_x(\varphi) < \infty$ in the sequel, and thus φ is in H_x.

For any $\varepsilon > 0$, let

$$\mathscr{T}_\varepsilon(\varphi) = \left\{ t \in [0, T]\colon \inf_{1 \leqslant i \leqslant n} \{\varphi_i(t), 1 - \varphi_i(t)\} > \varepsilon \right\} = \bigcup_{j=1}^{\infty} (\alpha_j, \beta_j).$$

By an argument similar to that used in the proof of (B) of Lemma 4.1, one can apply the monotone convergence theorem and get

$$I_x(\varphi) = \lim_{\varepsilon \to 0} \frac{1}{2} \int_{t \in \mathscr{T}_\varepsilon(\varphi)} \sum_{i=1}^n \frac{(\dot\varphi(t) - b_i(\varphi(t)))^2}{\varphi_i(t)} \, \mathrm{d}t,$$

and so (4.10) follows.

Next assume that x is in $\mathscr{S}_n^\circ \cap \mathscr{X}_p$. For convenience we assume that $p_1 > 0$. Since $I_x(\varphi)$ is finite, by applying the results in the proof of Lemma 4.1, $\varphi_1(t)$ will not hit zero at a later time. Thus $d = \inf_{0 \leqslant t \leqslant T} \varphi_1(t) > 0$. For any $\delta > 0$, choose ε small such that $(n-1)\varepsilon < \min\{\delta/2, d/2, x_i, i = 1, \ldots, x_n\}$. For any $0 \leqslant t \leqslant T$ and $i = 2, \ldots, n$, set $h_i^\varepsilon(t) = 0$ for $\varphi_i(t) > \varepsilon$; $h_i^\varepsilon(t) = \varepsilon - \varphi_i(t)$ for $\varphi_i(t) \leqslant \varepsilon$. Let

$$h_1^\varepsilon(t) = \sum_{i=2}^n h_i^\varepsilon(t), \quad \varphi^\varepsilon(t) = \varphi(t) + (-h_1^\varepsilon(t), h_2^\varepsilon(t), \ldots, h_{n-1}^\varepsilon(t)).$$

Then it is clear that $\varphi^\varepsilon(0) = \varphi(0) = x$, and $\varphi^\varepsilon(t) \in \mathscr{S}_n^\circ$ for t in $[0, T]$. It is also not hard to see that $\varphi_i^\varepsilon(t) = \varepsilon$ if $\varphi(t) \leqslant \varepsilon$. Note that for any $i = 2, \ldots, n$, if $p_i > 0$, then $h_i^\varepsilon(t) \equiv 0$ for small enough ε. Let $K_p = \{i \in \{2, \ldots, n\}\colon p_i = 0\}$.

By direct calculation, one gets

$$\begin{aligned} \liminf_{\gamma \to 0} \gamma \log P_x^\gamma\{B^\circ(\varphi, \delta; 0, T)\} &\geqslant \liminf_{\gamma \to 0} \gamma \log P_x^\gamma\{B^\circ(\varphi^\varepsilon, \delta/2; 0, T)\} \\ &\geqslant \lim_{\tilde\delta \to 0} \liminf_{\gamma \to 0} \gamma \log P_x^\gamma\{B^\circ(\varphi^\varepsilon, \tilde\delta/2; 0, T)\} \geqslant -I_x(\varphi^\varepsilon). \end{aligned} \tag{4.13}$$

Observe that

$$\begin{aligned} I_x(\varphi^\varepsilon) &= \frac{1}{2} \sum_{i=1}^n \int_0^T \frac{(\dot\varphi_i^\varepsilon(t) - b_i(\varphi^\varepsilon(t)))^2}{\varphi_i^\varepsilon(t)} \, \mathrm{d}t \\ &= \frac{1}{2} \left[\sum_{i \notin K_p \cup \{1\}} \int_0^T \frac{(\dot\varphi_i^\varepsilon(t) - b_i(\varphi^\varepsilon(t)))^2}{\varphi_i^\varepsilon(t)} \, \mathrm{d}t + \int_0^T \frac{(\dot\varphi_1^\varepsilon(t) - b_1(\varphi^\varepsilon(t)))^2}{\varphi_1^\varepsilon(t)} \, \mathrm{d}t \right. \\ &\quad \left. + \sum_{i \in K_p} \int_0^T \frac{(\dot\varphi_i^\varepsilon(t) - b_i(\varphi^\varepsilon(t)))^2}{\varphi_i^\varepsilon(t)} \, \mathrm{d}t \right] \end{aligned}$$

$$\leqslant \frac{1}{2}\left[\sum_{i\neq 1, i\notin K_p}\int_0^T \frac{(\dot{\varphi}_i^\varepsilon(t)-b_i(\varphi^\varepsilon(t)))^2}{\varphi_i^\varepsilon(t)}\,\mathrm{d}t + \int_0^T \frac{(\dot{\varphi}_1^\varepsilon(t)-b_1(\varphi^\varepsilon(t)))^2}{\varphi_1^\varepsilon(t)}\,\mathrm{d}t\right.$$

$$\left. + \sum_{i\in K_p}\int_0^T \frac{(\dot{\varphi}_i(t)-b_i(\varphi(t)))^2}{\varphi_i(t)}\,\mathrm{d}t + \frac{nT\theta^2}{4}\varepsilon\right], \tag{4.14}$$

where in the last inequality we used the fact that $\varphi_i^\varepsilon(t)=\varepsilon$ if $\varphi(t)\leqslant\varepsilon$. For any $i\notin K_{\mathrm{p}}$ and any t, we have $\varphi_i^\varepsilon(t)=\varphi(t)$ for small enough ε. Hence by letting ε go to zero, we have

$$\sum_{i\neq 1, i\notin K_p}\int_0^T \frac{(\dot{\varphi}_i^\varepsilon(t)-b_i(\varphi^\varepsilon(t)))^2}{\varphi_i^\varepsilon(t)}\,\mathrm{d}t \to \sum_{i\neq 1, i\notin K_p}\int_0^T \frac{(\dot{\varphi}_i(t)-b_i(\varphi(t)))^2}{\varphi_i(t)}\,\mathrm{d}t. \tag{4.15}$$

Let $\mathcal{N}_p=\{t\in[0,T]: \varphi_i(t)=0 \text{ for some } i\in K_p\}$. Noting that

$$(\dot{\varphi}_1^\varepsilon(t))^2\leqslant n\sum_{i\in K_p\cup\{1\}}\dot{\varphi}_i(t)^2. \tag{4.16}$$

The finiteness of $I_x(\varphi(\cdot))$ implies that $I_x^{\mathcal{N}_p}(\varphi(\cdot))=0$. By the dominated convergence theorem, we get that

$$\lim_{\varepsilon\to 0}\int_0^T \frac{(\dot{\varphi}_1^\varepsilon(t)-b_1(\varphi^\varepsilon(t)))^2}{\varphi_1^\varepsilon(t)}\,\mathrm{d}t = \lim_{\varepsilon\to 0}\left(\int_{\mathcal{N}_p^{\mathrm{c}}}+\int_{\mathcal{N}_p}\right)\frac{(\dot{\varphi}_1^\varepsilon(t)-b_1(\varphi^\varepsilon(t)))^2}{\varphi_1^\varepsilon(t)}\,\mathrm{d}t$$

$$=\int_0^T \frac{(\dot{\varphi}_1(t)-b_1(\varphi(t)))^2}{\varphi_1(t)}\,\mathrm{d}t. \tag{4.17}$$

This, combined with (4.14) and (4.15), implies

$$\limsup_{\varepsilon\to 0} I_x(\varphi^\varepsilon)\leqslant I_x(\varphi). \tag{4.18}$$

Finally, by letting ε go to zero, then δ go to zero in (4.13), we get (4.11). □

Remark. More detailed information about the way that a path leaves or approaches the boundary can be obtained by an argument similar to that used in Feng (2000), where the one-dimensional continuous branching processes are studied.

Let $\mathscr{Y}_p=\{x\in\mathscr{S}_n\colon x_i>0 \text{ whenever } p_i>0\}$. Then we get the following.

Theorem 4.3. *For any $x\in\mathscr{Y}_n$, the family $\{P_x^\gamma\}_{\gamma>0}$ satisfies a LDP on space $C([0,T],\mathscr{S}_n)$ with speed γ and good rate function $I_{x,p}(\cdot)$ given by*

$$I_{x,p}(\varphi)=\begin{cases}\dfrac{1}{2}\displaystyle\int_0^T\sum_{i=1}^n\frac{(\dot{\varphi}_i(t)-b_i(\varphi(t)))^2}{\varphi_i(t)}\,\mathrm{d}t, & \varphi\in H_{x,p},\\ \infty, & \varphi\notin H_{x,p},\end{cases} \tag{4.19}$$

where

$$H_{x,p}=\left\{\varphi\in C([0,T],\mathscr{S}_n)\colon \varphi(t)=x+\int_\alpha^t\dot{\varphi}(s)\,\mathrm{d}s,\ t\in[0,T]\right.$$

$$\left.\text{and } \varphi_i(t)\equiv 0 \text{ if } x_i=p_i=0\right\}.$$

Proof. If $x_i = p_i = 0$ for some $i = 1,\dots,n-1$, then $x_i(t) \equiv 0$ for all t in $[0,T]$. This explains the definition of the set $H_{x,p}$. By projection to lower dimension, the result is reduced to the case where $x_i + p_i > 0$ for all i. Then by applying Lemma 4.2 and Theorem 2.1, we get the result. □

Remark. This result has been proved in Theorem 3.3 in Dawson and Feng (1998) under the assumption that $x_i > 0, p_i > 0$ for all $i = 1,\dots,n$. Here we removed all restrictions on x and p in the upper bound, and extend the lower bound to cases where p_i can be zero for some $i = 1,\dots,n$.

5. LDP for FV processes

Path level LDPs are established in this section for FV processes with neutral mutation, and with selection.

5.1. LDP for FV processes with neutral mutation

Let $C^{1,0}([0,T]\times E)$ denote the set of all continuous functions on $[0,T]\times E$ with continuous first order derivative in time t. For any $\mu \in M_1(E)$, $\mu(\cdot)\in C([0,T],M_1(E))$, define

$$\mathscr{Y}_{v_0} = \{\mu \in M_1(E)\colon\ \mathrm{supp}(v_0)\subset \mathrm{supp}(\mu)\} \tag{5.1}$$

and

$$\begin{aligned} &S_\mu(\mu(\cdot))\\ &\quad= \sup_{g\in C^{1,0}([0,T])\times E)} \left\{ \langle \mu(T), g(T)\rangle - \langle \mu(0), g(0)\rangle - \int_0^T \langle \mu(s), \left(\frac{\partial}{\partial s} + A\right) g(s)\rangle\, \mathrm{d}s \right.\\ &\qquad \left. -\frac{1}{2}\int_0^T \int\int g(s,x)\,g(s,y)\,Q(\mu(s);\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s \right\}. \end{aligned} \tag{5.2}$$

Recall that for any μ in $M_1(E)$, we have $Q(\mu;\mathrm{d}x,\mathrm{d}y) = \mu(\mathrm{d}x)\delta_x(\mathrm{d}y) - \mu(\mathrm{d}x)\mu(\mathrm{d}y)$.

Let $B_{\mathrm{b}}^{1,0}([0,T]\times E)$ be the set all bounded measurable functions on $[0,T]\times E$ with continuous first-order derivative in time t, and stepwise constant in x. For every function f in $C^{1,0}([0,T]\times E)$ and any n, let $f_n(t,x) = f(t,i/n)$ for x in $[i/n,(i+1)/n)$, $f_n(t,1) = f_n(t,(n-1/n))$. Then it is clear that f_n is in $B_{\mathrm{b}}^{1,0}([0,T]\times E)$ and f_n converges to f uniformly as n goes to infinity. On the other hand, by interpolation, every element in $B_{\mathrm{b}}^{1}([0,T]\times E)$ can be approximated almost surely by a sequence in $C^{1,0}([0,T]\times E)$. Hence, we get

$$\begin{aligned} S_\mu(\mu(\cdot)) = \sup_{g\in B^{1,0}([0,T])\times E)} &\left\{ \langle \mu(T), g(T)\rangle - \langle \mu(0), g(0)\rangle \right.\\ &- \int_0^T \langle \mu(s), \left(\frac{\partial}{\partial s} + A\right) \mathrm{g}(s)\rangle\, \mathrm{d}s\\ &\left. -\frac{1}{2}\int_0^T \int\int g(s,x)\,g(s,y)\,Q(\mu(s);\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s \right\}. \end{aligned} \tag{5.3}$$

Definition 5.1. Let $C^\infty(E)$ denote the set of all continuous functions on E possessing continuous derivatives of all order. An element $\mu(\cdot)$ in $C([0,T],M_1(E))$ is said to be absolutely continuous as a distribution-valued function if there exist $M>0$ and an absolutely continuous function $h_M:[0,T]\to R$ such that for all $t,s\in[0,T]$

$$\sup_{\sup_{x\in E}|f(x)|<M} |\langle \mu(t),f\rangle - \langle \mu(s),f\rangle| \leqslant |h_M(t)-h_M(s)|.$$

Let $\mathscr{H}_\mu$ be the collection of all absolutely continuous paths in $C([0,T],M_1(E))$ starting at μ, and define

$$K_\mu(\mu(\cdot)) = \begin{cases} \int_0^T \|\dot\mu(s)-A^*(\mu(s))\|^2_{\mu(s)}\,\mathrm{d}s & \text{if } \mu(\cdot)\in\mathscr{H}_\mu,\\ \infty & \text{elsewhere,}\end{cases}$$

where A^* is the formal adjoint of A defined through the equality $\langle A^*(\mu),f\rangle = \langle \mu, Af\rangle$, and for any linear functional ϑ on space $C^\infty(E)$

$$\|\vartheta\|^2_\mu = \sup_{f\in C^\infty(E)}\left[\langle\vartheta,f\rangle - \frac{1}{2}\int_E\int_E f(x)f(y)Q(\mu;\mathrm{d}x,\mathrm{d}y)\right].$$

Then we have the following.

Theorem 5.1. *Assume $\mu\in\mathscr{Y}_{\nu_0}$. Then for any $\mu(\cdot)$ in $C([0,T],M_1(E))$, we have*

$$S_\mu(\mu(\cdot)) = K_\mu(\mu(\cdot)). \tag{5.4}$$

Proof. By definition, we have

$$S_\mu(\mu(\cdot))\leqslant K_\mu(\mu(\cdot)).$$

To get equality, we only need to consider the case when $S_\mu(\mu(\cdot))$ is finite. For any $[s,t]\subset[0,T]$, and any f in $C^{1,0}([s,t]\times E)$, let

$$l_{s,t}(f) = \langle \mu(t),f(t)\rangle - \langle\mu(s),f(s)\rangle - \int_s^t \langle \mu(u), \left(\frac{\partial}{\partial u}+A\right)f\rangle\,\mathrm{d}u.$$

By introducing an appropriate Hilbert space structure and applying the Riesz Representation Theorem, one can find a square integrable function h such that

$$l_{s,t}(f) = \int_s^t\int_E\int_E f(u,x)h(u,y)Q(\mu(u);\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}u, \tag{5.5}$$

$$\begin{aligned}\Gamma(f) &= \int_s^t\int_E\int_E \inf_{f\in C^{1,0}([s,t]\times E)}(h(u,x)-f(u,x))(h(u,y)\\ &\qquad -f(u,y))Q(\mu(u);\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}u\\ &\leqslant \inf_{f\in C^{1,0}([s,t]\times E)}\int_s^t\int_E\int_E(h(u,x)-f(u,x))(h(u,y)\\ &\qquad -f(u,y))Q(\mu(u);\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}u\\ &= 0\end{aligned} \tag{5.6}$$

and

$$S_\mu(\mu(\cdot)) = \frac{1}{2}\int_0^T \int_E \int_E h(u,x)h(u,y)Q(\mu(u);\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}u. \tag{5.7}$$

From (5.5), we can see that $\mu(\cdot)$ is absolutely continuous as a distribution-valued function. Applying (5.6),(5.7), and the fact that $\Gamma(f)\geqslant 0$, we get

$$\begin{aligned}\int_0^T \|\dot\mu(u) - A^*(\mu(u))\|_{\mu(u)}\,\mathrm{d}u &= \int_0^T \sup_{g\in C^\infty(E)} \{\langle \dot\mu(u) - A^*(\mu(u)), g\rangle \\ &\quad -\frac{1}{2}\int_E\int_E g(x)g(y)Q(\mu(u);\mathrm{d}x,\mathrm{d}y)\}\mathrm{d}u \\ &= \int_0^T \sup_{f\in C^{1,0}([0,T]\times E)} \{\langle\dot\mu(u), f(u)\rangle - \langle \mu(u), \dot f + Af\rangle \\ &\quad -\frac{1}{2}\int_E\int_E f(u,x)f(u,y)Q(\mu(u);\mathrm{d}x,\mathrm{d}y)\}\mathrm{d}u \\ &= S_\mu(\mu(\cdot)) - \tfrac{1}{2}\Gamma(f) = S_\mu(\mu(\cdot)).\end{aligned}$$

For further details, please refer to the appendix in Dawson and Feng (1998). □

Remark. If $E=\{1,\dots,n\}$ for some $n\geqslant 2$, then we have $C^\infty(E)=C(E)$, and the absolute continuity of $\mu(\cdot)$ as a distribution-valued function is the same as the usual absolute continuity as a real-valued function.

For μ in $\mathscr{Y}_{v_0}$, $\mu(\cdot)$ in $\mathscr{H}_\mu$ and $i=1,\dots,n$, let

$$\varphi(s) = (\varphi_1(s),\dots,\varphi_n(s)), \quad \varphi_i(s) = \mu(s,i), \quad b_i(\varphi(s)) = \theta/2(v_0(i) - \varphi_i(s)),$$

$$x_i = \varphi_i(0).$$

Then we have

$$\begin{aligned}\|\dot\mu(s) - A^*(\mu(s))\|^2_{\mu(s)} = \sup_{f\in C(E)} \Bigg[&\sum_{i=1}^n f(i)(\dot\varphi_i(s) - b_i(\varphi(s))) \\ &-\frac{1}{2}\sum_{i,j=1}^n f(i)f(j)\varphi_i(s)(\delta_{ij} - \varphi_j(s))\Bigg].\end{aligned}$$

If $\varphi_k(s) = \dot\varphi_k(s) - b_k(\varphi(s)) = 0$, then we have

$$\begin{aligned}\|\dot\mu(s) - A^*(\mu(s))\|^2_{\mu(s)} = \sup_{f\in C(E)} \Bigg[&\sum_{i\neq k} f(i)(\dot\varphi_i(s) - b_i(\varphi(s))) \\ &-\frac{1}{2}\sum_{i,j\neq k}^n f(i)f(j)\varphi_i(s)(\delta_{ij} - \varphi_j(s))\Bigg].\end{aligned}$$

This means that $\varphi_k(s)$ makes no contribution to $\|\dot\mu(s) - A^*(\mu(s))\|^2_{\mu(s)}$.

If $\varphi_k(s)=0, \dot{\varphi}_k(s)-b_k(\varphi(s))\neq 0$, then by choosing $f(i)=0$ for $i\neq k$, $f(k)=n\,\mathrm{sign}(\dot{\varphi}_k(s)-b_k(\varphi(s)))$, and let n go to infinity, we get that $\|\dot{\mu}(s)-A^*(\mu(s))\|^2_{\mu(s)}=\infty$.

Assume that $\varphi_i(s)>0$ for all $1\leqslant i\leqslant n$. Noting that $\sum_{i=1}^n b_i(\varphi(s))=0$, and $\sum_{i=1}^n \varphi_i(s)=1$, we get

$$\|\dot{\mu}(s)-A^*(\mu(s))\|^2_{\mu(s)} = \sup_{f\in C(E)}\left[\sum_{i=1}^{n-1}(f(i)-f(n))(\dot{\varphi}_i(s)-b_i(\varphi(s))) - \frac{1}{2}\sum_{i,j=1}^{n-1}(f(i)-f(n))(f(j)-f(n))\varphi_i(s)(\delta_{ij}-\varphi_j(s))\right].$$

Since the matrix $(\varphi_i(s)(\delta_{ij}-\varphi_j(s)))_{1\leqslant i,j\leqslant n-1}$ is invertible and its inverse is given by Eq. (3.4) in Dawson and Feng (1998), we get that

$$\|\dot{\mu}(s)-A^*(\mu(s))\|^2_{\mu(s)} = \sum_{i=1}^{n}\frac{(\dot{\varphi}_i(s)-b_i(\varphi(s)))^2}{\varphi_i(s)}.$$

Since we treat $0/0$ as zero, $c/0$ as infinity for $c>0$ in (4.2), we get that for finite type model,

$$I_{x,p}(\varphi)=S_\mu(\mu(\cdot)), \tag{5.8}$$

where $p=(v_0(1),\ldots,v_0(n))$. We thus derive a variational formula for the rate function obtained in Theorem 4.3.

For any $\mu\in M_1(E)$, let $P_\mu^{\theta,\gamma,v_0}$ be the law of the FV process with neutral mutation on space $C([0,T],M_1(E))$ starting at μ.

Lemma 5.2. *The family* $\{P_\mu^{\theta,\gamma,v_0}\}_{\gamma>0}$ *is exponentially tight, i.e., for any* $a>1$, *there exists a compact subset* $\mathscr{K}_a$ *of* $C([0,T],M_1(E))$ *such that*

$$\limsup_{\gamma\to 0}\gamma\log P_\mu^{\theta,\gamma,v_0}\{\mathscr{K}_a^{\mathrm{c}}\}\leqslant -a, \tag{5.9}$$

where $\mathscr{K}_a^{\mathrm{c}}$ *is the complement of* $\mathscr{K}_a$.

Proof. Let A be the neutral mutation operator. For any f in $C(E)$, let

$$M_t(f)=\langle\mu(t),f\rangle-\langle\mu(0),f\rangle-\int_0^t\langle\mu(s),Af\rangle\,\mathrm{d}s.$$

Then $M_t(f)$ is a martingale with increasing process

$$\langle\langle M(f)\rangle\rangle_t=\gamma\int_0^t\int\int f(x)f(y)Q(\mu(s);\mathrm{d}x,\mathrm{d}y)\mathrm{d}s.$$

More generally, let $M(\mathrm{d}t,\mathrm{d}x)$ denote the martingale measure obtained from $M_t(f)$. In other words, $M(\mathrm{d}t,\mathrm{d}x)$ is a martingale measure such that for any f in $C(E)$,

$$\int_0^t\int_E f(x)M(\mathrm{d}t,\mathrm{d}x)=M_t(f).$$

For any t in $[0,T]$, let $0=t_0<t_1<\cdots<t_m=t$ be any partition of $[0,t]$. For any function g in $C^{1,0}([0,T]\times E)$ and any u,v in $[0,T]$, let

$$M^{u,v}(g)=\langle\mu(v),g(v)\rangle-\langle\mu(u),g(u)\rangle-\int_u^v\left\langle\mu(s),\left(\frac{\partial}{\partial s}+A\right)g\right\rangle\mathrm{d}s.$$

Then we have

$$
\begin{aligned}
M^{0,t}(g) &= \sum_{k=1}^{m} M^{t_{k-1},t_k}(g)\\
&= \sum_{k=1}^{m}\Bigg[\langle \mu(t_k), g(t_k)-g(t_{k-1})\rangle + \langle \mu(t_k)-\mu(t_{k-1}), g(t_{k-1})\rangle\\
&\quad - \int_{t_{k-1}}^{t_k}\left\langle \mu(s), \left(\frac{\partial}{\partial s}+A\right) g\right\rangle \mathrm{d}s\Bigg]\\
&= \sum_{k=1}^{m}\Bigg[\langle \mu(t_k), g(t_k)-g(t_{k-1})\rangle - \int_{t_{k-1}}^{t_k}\left\langle \mu(s), \frac{\partial}{\partial s} g\right\rangle \mathrm{d}s\\
&\quad + \langle \mu(t_k)-\mu(t_{k-1}), g(t_{k-1})\rangle - \int_{t_{k-1}}^{t_k}\langle \mu(s), A[g(t_{k-1})+(g-g(t_{k-1}))]\rangle \mathrm{d}s\Bigg]\\
&= \sum_{k=1}^{m}\Bigg[\langle \mu(t_k), g(t_k)-g(t_{k-1})\rangle - \int_{t_{k-1}}^{t_k}\left\langle \mu(s), \frac{\partial}{\partial s} g\right\rangle \mathrm{d}s\Bigg]\\
&\quad + \sum_{k=1}^{m}\int_{t_{k-1}}^{t_k} g(t_{k-1}))M(\mathrm{d}s,\mathrm{d}x) - \sum_{k=1}^{m}\int_{t_{k-1}}^{t_k}\langle \mu(s), A(g-g(t_{k-1}))\rangle \mathrm{d}s.
\end{aligned}
$$

Letting $\max_{1\leqslant k\leqslant m}\{|t_k-t_{k-1}|\}$ go to zero, we get

$$
\begin{aligned}
\int_0^t\int_E g(s,x)M(\mathrm{d}s,\mathrm{d}x) &= \langle \mu(t), g(t)\rangle - \langle \mu(0), g(0)\rangle\\
&\quad - \int_0^t \left\langle \mu(s), \left(\frac{\partial}{\partial s}+A\right) g\right\rangle \mathrm{d}s, \qquad (5.10)
\end{aligned}
$$

which is a $P_\mu^{\theta,\gamma,v_0}$-martingale with increasing process

$$
\gamma\int_0^t\int\int g(s,x)g(s,y)Q(\mu(s);\mathrm{d}x,\mathrm{d}y)\mathrm{d}s.
$$

Hence by the exponential formula, for any real number α, the following

$$
Z_t^\gamma(g,\alpha)=\exp\left(\alpha\int_0^t\int g(s,x)M(\mathrm{d}s,\mathrm{d}x)-\frac{\gamma\alpha^2}{2}\int_0^t\int\int g(s,x)g(s,y)Q(\mu(s);\mathrm{d}x,\mathrm{d}y)\right)
$$

is a $P_\mu^{\theta,\gamma,v_0}$-martingale.

For any $\varepsilon>0, \delta\in[0,T/2)$, we have

$$
\sup_{\mu\in M_1(E)} P_\mu^{\theta,\gamma,v_0}\left\{\sup_{0\leqslant s<t\leqslant T, t-s<\delta}|\langle \mu(t), f\rangle - \langle \mu(s), f\rangle| > \varepsilon\right\} \qquad (5.11)
$$

$$\leqslant \sum_{k=0}^{[T/\delta]-1} P_{\mu}^{\theta,\gamma,v_0}\left\{\sup_{k\delta\leqslant s<t\leqslant(k+2)\delta\wedge T} |\langle \mu(t), f\rangle - \langle \mu(k\delta), f\rangle| > \frac{\varepsilon}{2}\right\} \tag{5.12}$$

$$\leqslant \frac{T}{\delta} \sup_{\mu\in M_1(E)} P_{\mu}^{\theta,\gamma,v_0}\left\{\sup_{t\in[0,2\delta)} |\langle \mu(t), f\rangle - \langle \mu(0), f\rangle| > \frac{\varepsilon}{2}\right\},$$

where the Markov property is used in the last inequality. For any fixed f in $C(E)$, the constant

$$c(f) = \sup_{\mu\in M_1(E)}\left[|\langle \mu, f\rangle| + \frac{1}{2}\int\int f(x)f(y)Q(\mu; \mathrm{d}x, \mathrm{d}y)\right]$$

is finite. Let $c(T, f) = c(f) \vee 2/T$. For any α we have

$$\alpha(\langle \mu(t), f\rangle - \langle \mu(0), f\rangle) - \alpha M_t(f) + \frac{\alpha^2}{2}\langle\langle M(f)\rangle\rangle_t$$

$$= \alpha\int_0^t \langle \mu(s), Af\rangle\,\mathrm{d}s + \frac{\alpha^2\gamma}{2}\int_0^t \int\int f(x)f(y)Q(\mu(s); \mathrm{d}x, \mathrm{d}y)\mathrm{d}s$$

$$\leqslant (1+\gamma\alpha)\alpha c(T, f)t,$$

which implies

$$\alpha(\langle \mu(t), f\rangle - \langle \mu(0), f\rangle) \leqslant (1+\gamma\alpha)\alpha c(T, f)t + \alpha M_t(f) - \frac{\alpha^2}{2}\langle\langle M(f)\rangle\rangle_t. \tag{5.13}$$

By Chebyshev's inequality and the martingale property, we get

$$P_{\mu}^{\theta,\gamma,v_0}\left\{\sup_{t\in[0,2\delta)} (\langle \mu(t), f\rangle - \langle \mu(0), f\rangle) > \frac{\varepsilon}{2}\right\}$$

$$= P_{\mu}^{\theta,\gamma,v_0}\left\{\sup_{t\in[0,2\delta)} Z_t^{\gamma}(f,\alpha) > \alpha\left(\frac{\varepsilon}{2} - 2(1+\gamma\alpha)c(T, f)\delta\right)\right\}$$

$$\leqslant \exp\left(-\alpha\left(\frac{\varepsilon}{2} - 2(1+\gamma\alpha)c(T, f)\delta\right)\right). \tag{5.14}$$

Choosing $\alpha = \beta/\gamma$ and minimizing with respect to $\beta \geqslant 0$ in (5.14), and by symmetry, one gets for $\varepsilon > 4c(T, f)\delta$

$$P_{\mu}^{\theta,\gamma,v_0}\left\{\sup_{t\in[0,2\delta)} |\langle \mu(t), f\rangle - \langle \mu(0), f\rangle| > \frac{\varepsilon}{2}\right\}$$

$$\leqslant 2\exp\left(-\frac{(\varepsilon - 4c(T, f)\delta)^2}{32c(T, f)\delta\gamma}\right). \tag{5.15}$$

For any $b > 1$, let

$$\delta_n = \frac{T}{2n^2}, \quad \varepsilon_n(b) = 9Tc(T, f)\sqrt{b/n}$$

and

$$\mathscr{K}_{f,b} = \bigcap_n \left\{\sup_{s<t\in[0,T],\, t-s<\delta_n} |\langle \mu(t), f\rangle - \langle \mu(s), f\rangle| \leqslant \varepsilon_n(b)\right\}.$$

Then we have

$$2(4c(T,f)\delta_n) = \frac{4Tc(T,f)}{n^2} < \frac{\varepsilon_n(b)}{2} \tag{5.16}$$

and

$$\frac{(\varepsilon_n(b) - 4c(T,f)\delta_n)^2}{32c(T,f)\delta_n} > \frac{81b(Tc(T,f))^2 n}{32Tc(T,f)} > 4bn. \tag{5.17}$$

By (5.15)–(5.17), we have

$$\begin{aligned}
&\sup_{\mu\in M_1(E)} P_\mu^{\theta,\gamma,\nu_0}\{\mathscr{K}_{f,b}^{\mathrm{c}}\} \\
&\quad\leqslant \sum_{n=1}^{\infty} \frac{T}{\delta_n} \sup_{\mu\in M_1(E)} P_\mu^{\theta,\gamma,\nu_0}\left\{\sup_{t\in[0,2\delta_n)} |\langle\mu(t),f\rangle - \langle\mu(0),f\rangle| > \frac{\varepsilon_n(b)}{2}\right\} \\
&\quad\leqslant \sum_{n=1}^{\infty} 4n^2 \exp\left(-\frac{4bn}{\gamma}\right) \leqslant 2\sum_{n=1}^{\infty} \exp\left(-\frac{2bn}{\gamma}\right) \leqslant 2\mathrm{e}^{-b/\gamma}.
\end{aligned} \tag{5.18}$$

Finally, choosing

$$\mathscr{K}_a = \bigcap_m \mathscr{K}_{f_m, ma},$$

and applying (5.18), we get (5.9). □

Lemma 5.3. *For any $\bar{\mu}(\cdot)$ in $C([0,T], M_1(E))$, we have*

$$\limsup_{\delta\to 0}\limsup_{\gamma\to 0} \gamma \log P_\mu^{\theta,\gamma,\nu_0}\{d(\mu(\cdot),\bar{\mu}(\cdot)) \leqslant \delta\} \leqslant -S_\mu(\bar{\mu}(\cdot)), \tag{5.19}$$

where

$$d(\mu(\cdot),\bar{\mu}(\cdot)) = \sup_{t\in[0,T]} \rho(\mu(t),\bar{\mu}(t)).$$

Proof. By definition, we have

$$S_\mu(\mu(\cdot)) = \sup_{g\in C^{1,0}([0,T])\times E)} \log Z_T^1(g,1)(\mu(\cdot)).$$

Applying Chebyshev's inequality, we get that for any $g \in C^{1,0}([0,T]\times E)$

$$\begin{aligned}
&P_\mu^{\theta,\gamma,\nu_0}\{d(\mu(\cdot),\bar{\mu}(\cdot)) \leqslant \delta\} \\
&\quad\leqslant \int \frac{Z_T^\gamma(\frac{1}{\gamma}g,1)(\mu(\cdot))}{\inf_{\{d(\mu(\cdot),\bar{\mu}(\cdot))\leqslant\delta\}} Z_T^\gamma((1/\gamma)g,1)(\mu(\cdot))}\,\mathrm{d}P_\mu^{\theta,\gamma,\nu_0} \\
&\quad\leqslant \left[\inf_{\{d(\mu(\cdot),\bar{\mu}(\cdot))\leqslant\delta\}} Z_T^\gamma\left(\frac{1}{\gamma}g,1\right)(\mu(\cdot))\right]^{-1},
\end{aligned}$$

which, combined with the relation $Z_T^\gamma(\frac{1}{\gamma}g,1) = (Z_T^1(g,1))^{1/\gamma}$, implies

$$\begin{aligned}
&\limsup_{\delta\to 0}\ \limsup_{\gamma\to 0} \gamma \log P_\mu^{\theta,\gamma,\nu_0}\{d(\mu(\cdot),\bar{\mu}(\cdot)) \leqslant \delta\} \\
&\quad\leqslant -\log Z_T^1(g,1)(\bar{\mu}(\cdot)).
\end{aligned} \tag{5.20}$$

The lemma follows by taking supremum with respect to g in (5.20). □

Lemma 5.4. *For any $\bar\mu(\cdot)$ in $C([0,T],M_1(E))$ and μ in $\mathscr{Y}_{v_0}$, we have*

$$\liminf_{\delta\to 0}\liminf_{\gamma\to 0}\gamma\log P_\mu^{\theta,\gamma,v_0}\{d(\mu(\cdot),\bar\mu(\cdot))<\delta\}\geqslant -S_\mu(\bar\mu(\cdot)). \tag{5.21}$$

Proof. If $S_\mu(\bar\mu(\cdot))=\infty$, the result is clear. Next we assume that $S_\mu(\bar\mu(\cdot))<\infty$. By Theorem 5.1, $\bar\mu(\cdot)$ is absolutely continuous in t as a distribution-valued function which implies the absolute continuity of $\mu(t,[a,b))$ and $\mu(t,[a,b])$ in t for $a,b\in E$.

Let $\{f_n\in C(E)\colon n\geqslant 1\}$ be a countable dense subset of $C(E)$. Define another metric $\hat d$ on $C([0,T],M_1(E))$ as follows:

$$\hat d(\mu(\cdot),v(\cdot))=\sup_{t\in[0,T]}\sum_{n=1}^\infty\frac{1}{2^n}(|\langle\mu(t),f_n\rangle-\langle v(t),f_n\rangle|\wedge 1). \tag{5.22}$$

Then d and $\hat d$ generate the same topology on $C([0,T],M_1(E))$. Hence it suffices to verify (5.21) with $\hat d$ in place of d. Clearly for any $\delta>0$, $\bar\mu\in C([0,T],M_1(E))$, there exists a $k\geqslant 1$ such that

$$\{\mu(\cdot)\in C([0,T],\,M_1(E))\colon|\langle\mu(t),f_n\rangle-\langle\bar\mu(t),f_n\rangle|\leqslant\delta/2,\,n=1,\ldots,k\}$$
$$\subset\{\hat d(\mu(\cdot),\bar\mu(\cdot))<\delta\}.$$

Next choose $x_0=0<x_1<x_2<\cdots<x_m<1=x_{m+1}$ such that

$$\max_{0\leqslant i\leqslant m,x,y\in[x_i,x_{i+1}]}\{|f_n(y)-f_n(x)|\colon 1\leqslant n\leqslant k\}\leqslant\frac{\delta}{6}.$$

Let

$$U_{x_1,\ldots,x_m}\left(\bar\mu(\cdot),\frac{\delta}{6\Gamma}\right)=\Bigg\{\mu(\cdot)\in C([0,T],\,M_1(E))\colon$$
$$\sup_{t\in[0,T],0\leqslant n\leqslant m-1}\{|\mu(t)([x_n,x_{n+1}))-\bar\mu(t)([x_n,x_{n+1}))|,$$
$$|\mu(t)([x_m,1])-\bar\mu(t)([x_m,1])|\}\leqslant\frac{\delta}{6\Gamma}\Bigg\},$$

where $\Gamma=\sup_{x,1\leqslant n\leqslant k}|f_n(x)|$. Then we have

$$U_{x_1,\ldots,x_m}\left(\bar\mu(\cdot),\frac{\delta}{6\Gamma}\right)\subset\{\hat d(\mu(\cdot),\bar\mu(\cdot))<\delta\}. \tag{5.23}$$

Set

$$\Theta(\mu(\cdot))=(\mu(\cdot)([0,x_1)),\ldots,\,\mu(\cdot)([x_m,1])).$$

The partition property (cf. Ethier and Kurtz, 1994) of the FV process implies that $P_{mu}^{\theta,\gamma,v_0}\circ\Theta^{-1}$ is the law of a FV process with finite type space. This combined with Theorem 4.3, implies

$$\liminf_{\delta\to 0}\liminf_{\gamma\to 0}\gamma\log P_{mu}^{\theta,\gamma,v_0}\{\hat d(\mu(\cdot),\bar\mu(\cdot))<\delta\}$$
$$\geqslant\liminf_{\delta\to 0}\liminf_{\gamma\to 0}\gamma\log P_v^{\theta,\gamma,v_0}\{U_{x_1,\ldots,x_m}(\bar\mu(\cdot),\frac{\delta}{6\Gamma})\}$$
$$\geqslant -I_{F(\mu),F(v_0)}(\Theta(\bar\mu))\geqslant -S_\mu(\bar\mu(\cdot)), \tag{5.24}$$

where the last inequality follows from Theorem 4.3, (5.8), and (5.3). □

Lemmas 5.2–5.4, combined with Theorem 2.1, imply the following.

Theorem 5.5. *For any μ in $\mathscr{Y}_{v_0}$, the family $P_\mu^{\theta,\gamma,v_0}$ satisfies a full LDP on $C([0,T], M_1(E))$ with a good rate function $S_\mu(\mu(\cdot))$.*

Remark. In Dawson and Feng (1998), the initial condition on μ is $\mathrm{supp}(\mu) = \mathrm{supp}(v_0)$. The current condition allows $\mathrm{supp}(v_0) \subset \mathrm{supp}(\mu)$.

5.2. LDP for FV processes with selection

For any $\mu \in M_1(E)$, let $P_\mu^{\theta,\gamma,V,v_0}$ be the law of the FV process on $C([0,T], M_1(E))$ with fitness function V and initial point μ. By the Cameron–Martin–Girsanov transformation (see Dawson, 1978) we have that,

$$\frac{\mathrm{d}P_\mu^{\theta,\gamma,V,v_0}}{\mathrm{d}P_\mu^{\theta,\gamma,v_0}} = Z_V(T) = \exp\left[\frac{1}{\gamma} G_V(\mu(\cdot))\right] > 0, \tag{5.25}$$

where

$$\begin{aligned} G_V(\mu(\cdot)) = {} & \int_0^T \int_E \left[\int_E V(y,z)\mu(s,\mathrm{d}z)\right] M(\mathrm{d}s,\mathrm{d}y) \\ & -\frac{1}{2}\int_0^T \int_E\int_E \left[\int_E V(x,z)\mu(s,\mathrm{d}z)\right] \\ & \times \left[\int_E V(y,z)\mu(s,\mathrm{d}z)\right] Q(\mu(s);\mathrm{d}x,\mathrm{d}y)\mathrm{d}s, \end{aligned} \tag{5.26}$$

and $M(\mathrm{d}s,\mathrm{d}y)$ is the same martingale measure as in (5.10). Define

$$R(\mu,\mathrm{d}x) = \int_E \left[\int_E V(y,z)\mu(\mathrm{d}z)\right] Q(\mu;\mathrm{d}x,\,\mathrm{d}y).$$

Then we have the following theorem.

Theorem 5.6. *For any $\mu \in \mathscr{Y}_{v_0}$, the family $\{P_\mu^{\theta,\gamma,V,v_0}\}$ satisfies a LDP on $C([0,T], M_1(E))$ as γ goes to zero with a good rate function $S_{\mu,V}(\mu(\cdot))$ given by*

$$\begin{aligned} S_{\mu,V}(\mu(\cdot)) &= S_\mu(\mu(\cdots)) - \Gamma_V(\mu(\cdot)) \\ &= \begin{cases} \displaystyle\int_0^T \|\dot{\mu}(s) - R(\mu(s)) - A^*(\mu(s))\|^2_{\mu(s)}\,\mathrm{d}s & \text{if } \mu(\cdot)\in\mathscr{H}_\mu, \\ \infty & \text{elsewhere}, \end{cases} \end{aligned} \tag{5.27}$$

where

$$\begin{aligned} \Gamma_V(\mu(\cdot)) = {} & \langle \mu(T), V(\mu(T))\rangle - \langle \mu(0), V(\mu(0))\rangle \\ & - \int_0^T \langle \mu(s), (\frac{\partial}{\partial s} + A)V(\mu(s))\rangle\,\mathrm{d}s \end{aligned}$$

$$-\frac{1}{2}\int_0^T \int_E \int_E \left[\int_E V(x,z)\mu(s,\mathrm{d}z)\right]\left[\int_E V(y,z)\mu(s,\mathrm{d}z)\right]$$
$$Q(\mu(s);\mathrm{d}x,\mathrm{d}y)\mathrm{d}s. \tag{5.28}$$

Proof. The identification of the two expressions of $S_{\mu,V}(\mu(\cdot))$ has been proved in Eq. (3.43) in Dawson and Feng (1998). Since $V(x)$ is continuous, we have that $\Gamma_V(\mu(\cdot))$ is bounded continuous on space $C([0,T],M_1(E))$. This, combined with the fact that $S_\mu(\mu(\cdot))$ is a good rate function, implies that $S_{\mu,V}(\mu(\cdot))$ is a good rate function. Let

$$m = \sup_{\mu(\cdot)\in C([0,T],M_1(E))}\left|\int_0^T \int_E \int_E \left[\int_E V(x,z)\mu(s,\mathrm{d}z)\right]\right.$$
$$\left.\times\left[\int_E V(y,z)\mu(s,\mathrm{d}z)\right] Q(\mu(s);\mathrm{d}x,\mathrm{d}y)\mathrm{d}s < \infty\right|. \tag{5.29}$$

For any measurable subset C of space $C([0,T],M_1(E))$, by using (5.25), Hölder's inequality, and martingle property, we get for any $\alpha>0,\beta>0,1/\alpha+1/\beta=1$,

$$P_\mu^{\theta,\gamma,V,v_0}\{C\}$$
$$=\int_C Z_V(T)\,\mathrm{d}P_\mu^{\theta,\gamma,v_0}$$
$$\leqslant \mathrm{e}^{m/2\gamma}\left(\int \exp\left[\frac{\alpha}{\gamma}\int_0^T\int_E\left(\int_E V(y,z)\mu(s,\mathrm{d}z)\right)M(\mathrm{d}s,\mathrm{d}y)\right]\mathrm{d}P_\mu^{\theta,\gamma,v_0}\right)^{1/\alpha} P_\mu^{\theta,\gamma,v_0}\{C\}^{1/\beta}$$
$$\leqslant \mathrm{e}^{(m/2\gamma)(1+\alpha)}P_\mu^{\theta,\gamma,v_0}\{C\}^{1/\beta}\left(\int Z_{\alpha V}(T)\mathrm{d}P_\mu^{\theta,\gamma,v_0}\right)^{1/\alpha}$$
$$=\mathrm{e}^{(m/2\gamma)(1+\alpha)}P_\mu^{\theta,\gamma,v_0}\{C\}^{1/\beta}, \tag{5.30}$$

which combined with the exponential tightness of the family $\{P_\mu^{\theta,\gamma,v_0}\}_{\gamma>0}$ implies that the family $\{P_\mu^{\theta,\gamma,V,v_0}\}_{\gamma>0}$ is also exponentially tight.

By choosing $C=\{v: d(v(\cdot),\mu(\cdot))\leqslant\delta\}$ in (5.30), we get

$$\lim_{\delta\to0}\limsup_{\gamma\to0}\gamma\log P_\mu^{\theta,\gamma,V,v_0}\{d(v(\cdot),\mu(\cdot))\leqslant\delta\}\leqslant\frac{m(1+\alpha)}{2}-\frac{1}{\beta}I_v(\mu(\cdot)) \tag{5.31}$$

which implies that for any $\mu(\cdot)$ in the complement of $\mathscr{H}_\mu$ one has

$$\lim_{\delta\to0}\liminf_{\gamma\to0}\gamma\log P_\mu^{\theta,\gamma,V,v_0}\{d(v(\cdot),\mu(\cdot))\leqslant\delta\}$$
$$=\lim_{\delta\to0}\limsup_{\gamma\to0}\gamma\log P_\mu^{\theta,\gamma,V,v_0}\{d(v(\cdot),\mu(\cdot))\leqslant\delta\}=-S_{\mu,V}(\mu(\cdot)). \tag{5.32}$$

Next we assume that $S_{\mu,V}(\mu(\cdot))<\infty$.

By an argument similar to that used in the derivation of (3.38) and (3.40) in Dawson and Feng (1998), we get that for any $\varepsilon>0$, there exists $\delta>0$ such that for any $\alpha>0,\beta>0,1/\alpha+1/\beta=1$,

$$\limsup_{\gamma\to0}\gamma\log P_\mu^{\theta,\gamma,V,v_0}\{d(v(\cdot),\mu(\cdot))\leqslant\delta\}\leqslant\Gamma_V(\mu(\cdot))+\frac{3+\alpha}{2}\varepsilon$$

$$+\frac{1}{\beta}\limsup_{\gamma\to 0}\gamma\log P_{\mu}^{\theta,\gamma,v_0}\{d(v(\cdot),\mu(\cdot))\leqslant\delta\}, \tag{5.33}$$

$$\liminf_{\gamma\to 0}\gamma\log P_{\mu}^{\theta,\gamma,V,v_0}\{d(v(\cdot),\mu(\cdot))<\delta\}\geqslant\Gamma_V(\mu(\cdot))-\left(2+\frac{\alpha}{2\beta}\right)\varepsilon$$

$$+\beta\liminf_{\gamma\to 0}\gamma\log P_{\mu}^{\theta,\gamma,v_0}\{d(v(\cdot),\mu(\cdot))<\delta\}. \tag{5.34}$$

Letting δ go to zero, then ε go to zero and finally β go to 1, we get that (5.32) is also true for $\mu(\cdot)$ satisfying $S_{\mu,V}(\mu(\cdot))<\infty$. Applying Theorem 2.1 again, we get the result. □

Acknowledgements

We thank the referee for helpful comments and suggestions, and W. Stannat for suggesting the type of example used in the proof of Lemma 3.3.

References

Dawson, D.A., 1978. Geostochastic calculus. Canad. J. Statist. 6, 143–168.

Dawson, D.A., Feng, S., 1998. Large deviations for the Fleming–Viot process with neutral mutation and selection. Stochastic Process. Appl. 77, 207–232.

Dawson, D.A., Gärtner, J., 1987. Large deviations from the McKean–Vlasov limit for weakly interacting diffusions. Stochastics 20, 247–308.

Dembo, A., Zeitouni, O., 1993. Large Deviations and Applications. Jones and Bartlett Publishers, Boston.

Donsker, M.D., Varadhan, S.R.S., 1975. Asymptotic evaluation of certain Markov process expectations for large time, I. Comm. Pure Appl. Math. 28, 1–47.

Ethier, S.N., Kurtz, T.G., 1993. Fleming–Viot processes in population genetics. SIAM J. Control Optim. 31 (2), 345–386.

Ethier, S.N., Kurtz, T.G., 1994. Convergence to Fleming–Viot processes in the weak atomic topology. Stochastic Process. Appl. 54, 1–27.

Feng, S., 2000. The behaviour near the boundary of some degenerate diffusions under random perturbation. In: Gorostiza, L.G., Gail Ivanoff, B., (eds.), Stochastic Models. CMS Conference Proceedings, Vol. 26, pp. 115–123.

Gänssler, P., 1971. Compactness and sequential compactness in spaces of measures. Z.Wahrsch, Verw. Geb. 17, 124–146.

Kingman, J.F.C., 1975. Random discrete distributions. J. Roy. Statist. Soc. Ser. 37, 1–20.

Pukhalskii, A.A., 1991. On functional principle of large deviations. In: Sazonov, V., Shervashidze, T. (Eds.), New Trends in Probability and Statistics, VSP Moks'las, Moskva.

Yosida, K., Hewitt, E., 1952. Finitely additive measures. Trans. Amer. Math. Soc. 72, 46–66.